

```
XX US5869291-A.
XX 09-FEB-1999.
XX 27-DEC-1996; 96US-0773423.
XX 27-DEC-1996; 96US-0773423.
XX (INCY-) INCYTE PHARM INC.
XX Bandman O, Goli SK, Hillman JL;
XX WPI; 1999-152769/13.
XX N-PSDB; AAX04755.
XX DNA encoding human Rab proteins - useful for producing recombinant
XX proteins
XX Example 1; Fig 3A-B; 36pp; English.
XX The present sequence represents Rab protein designated HRAB-3. The
XX nucleic acid sequence was first identified in Incyte clone 641412
XX from a breast tissue cDNA library. The HRAB nucleic acid sequences
XX are useful for producing recombinant human Rab proteins which may
XX be useful for treating disorders involving vesicle targeting,
XX membrane transfer or fusion, or protein processing, targeting or
XX secretion, e.g. disorders characterised by abnormal neurotransmitter
XX production and release, lysosomal storage diseases, chorioidaemia and
XX endocrine hypofunction and hyperfunction.
XX Sequence 201 AA;
Query Match 100.0%; Score 1062; DB 20; Length 201;
Best Local Similarity 100.0%; Pred. No. 3.8e-106;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGKSLFKVILLGGVGKSLMRYNTNKFDTQLTFTIGVEFLNKDLEVDGHFTVMTQI 60
DB 1 MAGKSLFKVILLGGVGKSLMRYNTNKFDTQLTFTIGVEFLNKDLEVDGHFTVMTQI 60
QY 61 WDTAGQERFSLRTPFYRGSDCLLTFSVDDSQSQFQNLNWKKEFIYYADVKEPSPFFV 120
DB 61 WDTAGQERFSLRTPFYRGSDCLLTFSVDDSQSQFQNLNWKKEFIYYADVKEPSPFFV 120
QY 121 ILGNKIDISERQVSTEEAQACRNDGYPYFETSADATNVAAAFEEAVRRVLATEDRSD 180
DB 121 ILGNKIDISERQVSTEEAQACRNDGYPYFETSADATNVAAAFEEAVRRVLATEDRSD 180
QY 181 HLIQDTVNLHRKPKPSSSCC 201
DB 181 HLIQDTVNLHRKPKPSSSCC 201
RESULT 2
ID ABP53078 standard; Protein; 201 AA.
XX ABP53078;
XX ABP53078;
XX 08-NOV-2002 (first entry)
XX Human Rab protein HRAB-3 SEQ ID NO:5.
XX Human; Rab protein; HRAB-3; antianaemic; antidiabetic; antiinflammatory;
XX immunosuppressive; dermatological; neuroprotective; antitumor; anti-HIV;
XX antirheumatoid; antiarthritic; cycostatic; endocrine; gene therapy;
XX immune system disorder; autoimmune haemolytic anaemia; myasthenia gravis;
XX juvenile diabetes mellitus; lupus erythematosus; multiple sclerosis;
XX Crohn's disease; rheumatoid arthritis; AIDS; ulcerative colitis; cancer;
XX leukaemia; lymphoma; carcinoma; vesicle targeting disease; ras family;
XX lysosomal storage disease; chorioidaemia; endocrine hypofunction;
XX endocrine hyperfunction.
```

```
XX OS Homo sapiens.
XX US2002090712-A1.
XX 11-JUL-2002.
XX 19-NOV-2001; 2001US-0988974.
XX 27-DEC-1996; 96US-0773423.
XX 17-DEC-1998; 98US-0215887.
XX (INCY-) INCYTE PHARM INC.
XX Hillman JL, Goli SK, Bandman O;
XX WPI; 2002-655836/70.
XX N-PSDB; ABQ75524.
XX Novel Rab polypeptides, low molecular weight guanidine triphosphatases
XX of Ras superfamily, useful for treating, preventing or diagnosing
XX cancer, and immune system disorders e.g. ulcer, arthritis and Crohn's
XX disease -
XX Claim 1; Fig 3A-B; 40pp; English.
XX The present sequence represents human HRAB-3 which is a Rab protein (low
XX molecular weight guanidine triphosphatases of the Ras superfamily).
XX Rab proteins (I) have antianaemic, antidiabetic, immunosuppressive,
XX antiinflammatory, dermatological, neuroprotective, antirheumatoid,
XX antiulcer, antiarthritic, anti-HIV, cytostatic and endocrine activities,
XX and can be used in gene therapy. (I) can be used for screening a compound
XX for effectiveness as an agonist or antagonist, for screening a compound
XX that specifically binds to (I) or modulates the activity of (I), and for
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.
XX (I) can be used for diagnosis, treating or preventing immune system
XX disorders (including autoimmune haemolytic anaemia, juvenile diabetes
XX mellitus, lupus erythematosus, multiple sclerosis, myasthenia gravis,
XX Crohn's disease, rheumatoid arthritis, AIDS or ulcerative colitis),
XX cancer (including leukaemia, lymphoma, carcinomas of lung, penis, breast,
XX pancreas, colon, stomach, small intestine, brain or prostate), diseases
XX involving vesicle targeting, membrane transfer or fusion, or protein
XX processing, targeting or secretion, and conditions characterised by
XX abnormal levels of neurotransmitter production and release, lysosomal
XX storage diseases, chorioidaemia, endocrine hypofunction and
XX hyperfunction.
XX Sequence 201 AA;
Query Match 100.0%; Score 1062; DB 23; Length 201;
Best Local Similarity 100.0%; Pred. No. 3.8e-106;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGKSLFKVILLGGVGKSLMRYNTNKFDTQLTFTIGVEFLNKDLEVDGHFTVMTQI 60
DB 1 MAGKSLFKVILLGGVGKSLMRYNTNKFDTQLTFTIGVEFLNKDLEVDGHFTVMTQI 60
QY 61 WDTAGQERFSLRTPFYRGSDCLLTFSVDDSQSQFQNLNWKKEFIYYADVKEPSPFFV 120
DB 61 WDTAGQERFSLRTPFYRGSDCLLTFSVDDSQSQFQNLNWKKEFIYYADVKEPSPFFV 120
QY 121 ILGNKIDISERQVSTEEAQACRNDGYPYFETSADATNVAAAFEEAVRRVLATEDRSD 180
DB 121 ILGNKIDISERQVSTEEAQACRNDGYPYFETSADATNVAAAFEEAVRRVLATEDRSD 180
QY 181 HLIQDTVNLHRKPKPSSSCC 201
DB 181 HLIQDTVNLHRKPKPSSSCC 201
RESULT 3
ABP31125
ID ABB31125 standard; Peptide; 209 AA.
```



Db 129 ILGNKIDISERQVSTEEAQCWCRNGDYPYFETSAKDATNVAAAFEEAVRVLATEDRSD 188  
QY 181 HLIQTDVTNLRKPKPSSSCC 201  
Db 189 HLIQTDVTNLRKPKPSSSCC 209  
RESULT 5  
AAM17322  
ID AAM17322 standard; Protein; 209 AA.  
XX  
AC AAM17322;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Peptide #3756 encoded by probe for measuring cervical gene expression.  
XX  
KW Probe; human; microarray; gene expression; cervical epithelial cell;  
KW cervical cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157278-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00670.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488901/53.  
XX  
DR Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -  
XX  
XX Claim 27; SEQ ID No 22148; 487pp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
CC (SENP: see AAI10068-AA128459). The present sequence is a peptide encoded  
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 209 AA;  
Query Match 100.0%; Score 1062; DB 22; Length 209;  
Best Local Similarity 100.0%; Pred. No. 4e-106;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAGKSLFKVILLGGVGKSLMRYVTNKFDTQLFHTIGVEFLNKDLVDGHHFTVMOI 60  
Db 9 MAGKSLFKVILLGGVGKSLMRYVTNKFDTQLFHTIGVEFLNKDLVDGHHFTVMOI 68  
QY 61 WDTAGQERFRSLRTPPYGSDCLLTFVSQDSQFQNLNWKKBFIYYADYKESPFV 120  
Db 69 WDTAGQERFRSLRTPPYGSDCLLTFVSQDSQFQNLNWKKBFIYYADYKESPFV 128

QY 121 ILGNKIDISERQVSTEEAQCWCRNGDYPYFETSAKDATNVAAAFEEAVRVLATEDRSD 180  
Db 129 ILGNKIDISERQVSTEEAQCWCRNGDYPYFETSAKDATNVAAAFEEAVRVLATEDRSD 188  
QY 181 HLIQTDVTNLRKPKPSSSCC 201  
Db 189 HLIQTDVTNLRKPKPSSSCC 209  
RESULT 6  
AAM05002  
ID AAM05002 standard; Protein; 209 AA.  
XX  
AC AAM05002;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Peptide #3684 encoded by probe for measuring breast gene expression.  
XX  
KW Probe; human; breast disease; breast cancer; development disorder;  
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
OS Homo sapiens.  
XX  
PN WO200157270-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 29-JAN-2001; 2001WO-US00661.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-476286/51.  
XX  
DR Novel single exon nucleic acid probe used to measuring gene expression  
PT in a human breast -  
XX  
XX Claim 27; SEQ ID No 13742; 322pp; English.  
XX  
XX The present invention relates to novel single exon nucleic acid probes  
CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for measuring human gene expression in  
CC a human breast sample, where the probe hybridises at high stringency to a  
CC nucleic acid expressed in the human breast. The probes are useful for  
CC predicting, diagnosing, grading, staging, monitoring and prognosis  
CC diseases of the human breast, particularly those diseases with polygenic  
CC aetiology. The diseases include: breast cancer; disorders of development,  
CC inflammatory diseases of the breast; fibrocystic changes; proliferative  
CC breast disease and non-carcinoma tumours.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 209 AA;  
Query Match 100.0%; Score 1062; DB 22; Length 209;  
Best Local Similarity 100.0%; Pred. No. 4e-106;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAGKSLFKVILLGGVGKSLMRYVTNKFDTQLFHTIGVEFLNKDLVDGHHFTVMOI 60  
Db 9 MAGKSLFKVILLGGVGKSLMRYVTNKFDTQLFHTIGVEFLNKDLVDGHHFTVMOI 68

QY 61 WDTAGQERFRSLRTPFYRGSDCLLTFSVDDSQSFQNLNNKKKEFYIYADVKEPESFPFV 120  
 DB 69 WDTAGQERFRSLRTPFYRGSDCLLTFSVDDSQSFQNLNNKKKEFYIYADVKEPESFPFV 128  
 QY 121 ILGNKIDISERQVSTEEAQAQCRDNGDYPYFETSADKATNVAAAFEEAVRRVLATEPDRSD 180  
 DB 129 ILGNKIDISERQVSTEEAQAQCRDNGDYPYFETSADKATNVAAAFEEAVRRVLATEPDRSD 188  
 QY 181 HLIQTDVNLHHRKPKPSSCC 201  
 DB 189 HLIQTDVNLHHRKPKPSSCC 209  
 RESULT 7  
 ID ABG39111 standard; Peptide; 209 AA.  
 AC ABG39111;  
 XX  
 DT 19-AUG-2002 (first entry)  
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 28776.  
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW Chronic obstructive pulmonary disease; interstitial lung disease;  
 KW Familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW Tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW Pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW Pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW Primary ciliary dyskinesia; pulmonary hypertension;  
 KW Hyaline membrane disease.  
 XX Homo sapiens.  
 OS  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00665.  
 XX  
 PR 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX  
 PS Claim 27; SEQ ID No 28776; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the labeled nucleic acids bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences

of the eukaryote; and (b) detecting specific hybridisation of detectably  
 labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,  
 having a fragment identical to the predicted exon, the probe is included  
 in the above mentioned microarray; assigning exons to a single gene,  
 comprising (a) identifying exons from genomic sequence by the method  
 above and (b) measuring the expression of each of the exons in several  
 tissues and/or cell types using hybridisation to a single exon  
 microarrays having a probe with the exon, where a common pattern of  
 expression of the exons in the tissues and/or cell types indicates that  
 the exons should be assigned to a single gene; a peptide comprising one  
 of 12011 sequences, mentioned in the specification, or encoded by the  
 probes/open reading frames (ORF). The probes are used for gene  
 expression analysis, and for identifying exons in a gene, particularly  
 using human lung derived mRNA and for the study of lung diseases  
 such as asthma, lung cancer, chronic obstructive pulmonary disease  
 (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 fibrosis, neurofibromatosis, tuberosus sclerosis, Gaucher's disease,  
 Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
 pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 and hyaline membrane disease. The present sequence is a peptide/protein  
 encoded by a single exon probe of the invention.  
 Note: The sequence data for this patent did not form part  
 of the printed specification, but was obtained in electronic  
 format directly from WIPO at  
 ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 209 AA;  
 Query Match 100.0%; Score 1062; DB 23; Length 209;  
 Best Local Similarity 100.0%; Pred. No. 4e-106;  
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAGKSLFKVILLGDGCVKSLMRYVTKKDTQLFHTIGVEFLNKDLVGDGHFTVMOI 60  
 DB 9 MAGKSLFKVILLGDGCVKSLMRYVTKKDTQLFHTIGVEFLNKDLVGDGHFTVMOI 68  
 QY 61 WDTAGQERFRSLRTPFYRGSDCLLTFSVDDSQSFQNLNNKKKEFYIYADVKEPESFPFV 120  
 DB 69 WDTAGQERFRSLRTPFYRGSDCLLTFSVDDSQSFQNLNNKKKEFYIYADVKEPESFPFV 128  
 QY 121 ILGNKIDISERQVSTEEAQAQCRDNGDYPYFETSADKATNVAAAFEEAVRRVLATEPDRSD 180  
 DB 129 ILGNKIDISERQVSTEEAQAQCRDNGDYPYFETSADKATNVAAAFEEAVRRVLATEPDRSD 188  
 QY 181 HLIQTDVNLHHRKPKPSSCC 201  
 DB 189 HLIQTDVNLHHRKPKPSSCC 209  
 RESULT 8  
 ABP53081  
 ID ABP53081 standard; Protein; 201 AA.  
 XX  
 AC ABP53081;  
 XX  
 DT 08-NOV-2002 (first entry)  
 XX  
 DE Canine Rab99 protein SEQ ID NO:9.  
 XX  
 KW Human; Rab protein; HRAS; antianemic; antidiabetic; antinflammatory;  
 KW immunosuppressive; dermatological; neuroprotective; antiulcer; anti-HIV;  
 KW antirheumatoid; antirheumatic; cytostatic; endocrine; gene therapy;  
 KW immune system disorder; autoimmune haemolytic anaemia; myasthenia gravis;  
 KW juvenile diabetes mellitus; lupus erythematosus; multiple sclerosis;  
 KW Crohn's disease; rheumatoid arthritis; AIDS; ulcerative colitis; cancer;  
 KW leukaemia; lymphoma; carcinoma; vesicle targeting disease; ras family;  
 KW lysosomal storage disease; chorioideraemia; endocrine hypofunction;  
 KW endocrine hyperfunction.  
 XX  
 OS Canis sp.  
 XX



PN US2002090712-A1.  
 XX 11-JUL-2002.  
 XX 19-NOV-2001; 2001US-0988974.  
 XX 27-DEC-1996; 96US-0773423.  
 XX 17-DEC-1998; 98US-0215887.  
 XX (INCY-) INCYTE PHARM INC.  
 XX Hillman JL, Goli SK, Bandman O;  
 XX WPI; 2002-655836/70.  
 XX Novel Rab polypeptides, low molecular weight guanidine triphosphatases  
 PT of Ras superfamily, useful for treating, preventing or diagnosing  
 PT cancer, and immune system disorders e.g. ulcer, arthritis and Crohn's  
 PT disease -  
 XX Disclosure; Fig 4; 40pp; English.  
 XX The present invention describes human Rab proteins (low molecular weight  
 CC guanidine triphosphatases of the Ras superfamily) designated HRABs.  
 CC Rab proteins (I) have antidiabetic, antidiabetic, immunosuppressive,  
 CC antiinflammatory, dermatological, neuroprotective, antirheumatoid,  
 CC antitumor, antiarthritic, anti-HIV, cytostatic and endocrine activities,  
 CC and can be used in gene therapy. (I) can be used for screening a compound  
 CC for effectiveness as an agonist or antagonist, for screening a compound  
 CC that specifically binds to (I) or modulates the activity of (I), and for  
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
 CC (I) can be used for diagnosing, treating or preventing immune system  
 CC disorders (including autoimmune haemolytic anaemia, juvenile diabetes  
 CC mellitus, lupus erythematosus, multiple sclerosis, myasthenia gravis,  
 CC Crohn's disease, rheumatoid arthritis, AIDS or ulcerative colitis),  
 CC cancer (including leukemia, lymphoma, carcinomas of lung, penis, breast,  
 CC pancreas, colon, stomach, small intestine, brain or prostate) diseases  
 CC involving vesicle targeting, membrane transfer or fusion, or protein  
 CC processing, targeting or secretion, and conditions characterised by  
 CC abnormal levels of neurotransmitter production and release, lysosomal  
 CC storage diseases, chorioideraemia, endocrine hypofunction and  
 CC hyperfunction. The present sequence represents a Rab protein given in  
 CC comparison with the human HRABs (HRABs-1, HRABs-2 and HRABs-3) from the  
 CC present invention.  
 XX SQ Sequence 201 AA;  
 Query Match 99.5%; Score 1057; DB 23; Length 201;  
 Best Local Similarity 99.5%; Pred. No. 1.3e-105;  
 Matches 200; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAGKSSLEKVLGGGVGKSSLMNRYVYTNKFDLTQFTHTIGVEFLNKDLEVDGHFTVMOI 60  
 DB 1 MAGKSSLEKVLGGGVGKSSLMNRYVYTNKFDLTQFTHTIGVEFLNKDLEVDGHFTVMOI 60  
 QY 61 WDTAQERFSLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPESPPFV 120  
 DB 61 WDTAQERFSLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPESPPFV 120  
 QY 121 ILGNKIDISERQVSTEEAQAACRONGDYPYPTSAKATNVAAAFEEAVRVLATEDRSD 180  
 DB 121 ILGNKIDISERQVSTEEAQAACRONGDYPYPTSAKATNVAAAFEEAVRVLATEDRSD 180  
 QY 181 HLIQDTVNLHRKPKPSSSCC 201  
 DB 181 HLIQDTVNLHRKPKPSSSCC 201  
 RESULT 9  
 AAB19172  
 ID AAB19172 standard; Protein; 201 AA.  
 XX  
 AC AAB19172;

XX 19-FEB-2001 (first entry)  
 XX Amino acid sequence of human RAB34.  
 XX RAB protein; GTPase; GTP binding; gene therapy; cancer.  
 XX Homo sapiens.  
 XX WO200058464-A2.  
 XX 05-OCT-2000.  
 XX 13-MAR-2000; 2000WO-US06330.  
 XX 25-MAR-1999; 99US-0126083.  
 XX (AXYS-) AXYS PHARM INC.  
 XX Allen M, Abel K, McIntosh B, Vega R, Rutter M, Buckler AJ;  
 XX WPI; 2000-647233/62.  
 XX N-PSDB; AA96894.  
 XX Novel isolated nucleic acid encoding a mammalian RAB protein useful for  
 PT identifying homologous or related genes, in producing composition that  
 PT modulates expression or function of RAB for cancer therapy -  
 XX Claim 2; Page 50-51; 58pp; English.  
 XX The present sequence represents a mammalian RAB protein. RAB proteins  
 CC constitute the largest family of small GTPases, with over 40 identified  
 CC isoforms. RAB proteins contain four highly conserved peptide sequences  
 CC involved in GTP binding and hydrolysis. Compositions comprising RAB  
 CC nucleic acid are useful for identifying homologous or related genes,  
 CC in producing compositions that modulate the expression or function of  
 CC RAB for gene therapy, mapping functional regions of the protein and  
 CC in studying associated physiological pathways. In addition, modulation  
 CC of the gene activity in vivo is used for prophylactic and therapeutic  
 CC purposes, such as treatment of cancer, and identification of cell type  
 CC based on expression. The DNA may also be used to identify expression of  
 CC the gene in a biological specimen.  
 XX SQ Sequence 201 AA;  
 Query Match 79.0%; Score 839; DB 21; Length 201;  
 Best Local Similarity 76.1%; Pred. No. 4.3e-82;  
 Matches 153; Conservative 22; Mismatches 26; Indels 0; Gaps 0;  
 QY 1 MAGKSSLEKVLGGGVGKSSLMNRYVYTNKFDLTQFTHTIGVEFLNKDLEVDGHFTVMOI 60  
 DB 1 MSGKSLLLKVLGGGVGKSSLMNRYVYTNKFDLTQFTHTIGVEFLNRDLVDGFRVLIQI 60  
 QY 61 WDTAQERFSLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPESPPFV 120  
 DB 61 WDTAQERFSLRTPFYRGADCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPESPPFV 120  
 QY 121 ILGNKIDISERQVSTEEAQAACRONGDYPYPTSAKATNVAAAFEEAVRVLATEDRSD 180  
 DB 121 VLGNKVDKEDQVTTTETEAQTWCMEGDIPLYLETSKODTNTVAFEEAVRQVLAEEQLE 180  
 QY 181 HLIQDTVNLHRKPKPSSSCC 201  
 DB 181 HCVLGHITDNLGSKAGSSCC 201  
 RESULT 10  
 ABP43693  
 ID ABP43693 standard; Protein; 201 AA.  
 XX  
 AC ABP43693;  
 XX  
 DT 26-FEB-2003 (first entry)



PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 08-SEP-2000; 2000US-0231968.  
 PR 12-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235835.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 13-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 20-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241788.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.

PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-465460/50.  
 DR N-PSDB; AAS27473.  
 XX  
 PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
 PT prognosing disorders related to the proteins, including cancers, immune  
 PT disorders and neuronal disorders -  
 XX  
 PS Claim 1; SEQ ID No 1121; 880pp; English.  
 XX  
 CC The invention relates to novel isolated polypeptides (I), and  
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
 CC diagnosing, preventing and treating diseases including immune system  
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
 CC transplant rejections and graft versus host disease, infectious diseases  
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative  
 CC disorders, primary haematopoietic disorders, hyperproliferative  
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal  
 CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal  
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders  
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in  
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
 CC Addison's disease), reproductive system disorders, gastrointestinal  
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),  
 CC as stimulators of B-cell responsiveness to pathogens, activators of  
 CC T-cells, to induce higher affinity antibodies, and as a means to induce  
 CC tumour proliferation in pathologies e.g. acquired immune deficiency  
 CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction  
 CC pathway protein, amino acid sequences of the invention.  
 XX

Query Match 79.0%; Score 839; DB 22; Length 222;

Best Local Similarity 76.1%; Pred. No. 4.9e-82;  
 Matches 153; Conservative 22; Mismatches 26; Indels 0; Gaps 0;

QY 1 MAGKSLFKVILLGGGVGKSSLMRYNTKPDTLFTTIGVEFLNKDLVDGHEVTTMOI 60  
 Db 22 MSGKSILLKVVILLGGGVGKSSLMRYNTKPDLSAFTTIGVEFLNRDLVDGHEVTTQI 81  
 QY 61 WDTAGQERPSRLTFFYRGSDCLLTFFSDVDSQSQNLNNKKEFIYADVKPEPSFPV 120  
 Db 82 WDTAGQERPSRLTFFYRGSDCLLTFFSDVDSQSQNLNNKKEFIYADVKPEPSFPV 141  
 QY 121 ILGNKIDISEROVSTEEAQWCRDNGDYPYPTSAKDATNVAAAFEEAVRRVLATEDRSD 180  
 Db 142 VLGNKVDKEDRQVTTTAAQTWCMNGDIPYLETSAKDNTVTVAEEAVRQVLAVEEQLE 201



PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235835.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465460/50.  
XX N-PSDB; AAS27054.  
PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
PT prognosing disorders related to the proteins, including cancers, immune  
PT disorders and neuronal disorders -  
XX Claim 1; SEQ ID No 702; 880pp; English.  
XX The invention relates to novel isolated polypeptides (I), and  
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
CC diagnosing, preventing and treating diseases including immune system  
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
CC transplant rejections and graft versus host disease, infectious diseases  
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
CC other blood-related disorders (sickle cell anaemia), myeloproliferative  
CC disorders, primary haematopoietic disorders, hyperproliferative  
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal  
CC disorders (e.g. glomerulonephritis), cardiovascular disorders  
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in  
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
CC Addison's disease), reproductive system disorders, gastrointestinal  
CC disorder (inflammatory disorders), liver disorders (cirrhosis),  
CC as stimulators of B-cell responsiveness to pathogens, activators of  
CC T-cells, to induce higher affinity antibodies, and as a means to induce  
CC tumour proliferation in pathologies e.g. acquired immune deficiency  
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction  
CC pathway protein, amino acid sequences of the invention.  
XX

Query Match 76.5%; Score 812.5; DB 22; Length 221;

```
Best Local Similarity 75.5%; Pred. No. 3.5e-79;
Matches: 151; Conservative 22; Mismatches 26; Indels 1; Gaps 1;
QY 1 MACKSIFKVLIGDGGVKGKSLMNRVYTNKPTQLPHYIGVEFLNKDLEVDGHHFTWQI 60
Db 22 MSGKSLLLKVLIGDGGVKGKSLMNRVYTNKPTQAPHTIGVEFLNRDLEVDGHHFTWQI 81
QY 61 WDTAGQERFSLRTPFYRGSDCCLLTFSVDDSQSFQNLNNWKKFTYYADVKPEPSFPFV 120
Db 82 WDTAGQERFSLRTPFYRGADCLLTFSVDDRSFENLGNWQKEFYIYADVKDPHPFV 141
QY 121 ILGNKIDISERQVSTEAQWCDNGDYPYFETSAKDQATNVAAPEAVRRVLATEDRSD 180
Db 142 VLGNKYDKEDRQVTTETQATWCMENGDPYLETSAKDDTNTVTAPEAVRQVLAVEEQLE 201
QY 181 HLIQTDTVNLHRKPKPSSSC 200
Db 202 HCMLGHTIDL-TGLQAGSSC 220

RESULT 14
AAGS4480
ID AAGS4480 standard; Protein; 201 AA.
XX AC AAGS4480;
XX DT 18-OCT-2000 (first entry)
XX DE Zea mays protein fragment SEQ ID NO: 69470.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence; corn.
XX OS Zea mays subsp. mays.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000BP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
```

```
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156559.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 50.4%; Score 535; DB 21; Length 201;
Best Local Similarity 51.5%; Pred. No. 2.6e-49;
Matches 102; Conservative 35; Mismatches 57; Indels 4; Gaps 2;

QY 7 LFKVILLGDGVGKSLMRYVTKEDTOLFTHTIGVEELNKDLEVDGHFVFWQIWDTAGQ 66
Db 2 LLKVIILGDSGVGKSLMRYVTKEDTOLFTHTIGVEELNKDLEVDGHFVFWQIWDTAGQ 61
QY 67 ERFRSLRTPFYRGSDCCLLTFSVDDSQFQNLNNKKKEFYIYADVKKPESPFPVILGNKI 126
Db 62 ERFQSLGVAFYRGADCCVLVYDVNKSKEKLNWREELIQASPSDPENFPFVILGNKI 121
QY 127 DI---SERQVSTEEAQWCEGDYFYFETSADKATNVAAPFEAVRRVLATEDRSD-HL 182
Db 122 DVDGNSRTVSEKKAKAWCASKGNIPYFETSABKGFVVEAFECTARNVAKNEPEEDMYL 181
QY 183 IQDTVNLHKKPKPSSSC 200
Db 182 PDTIDVGAGRQQRSSGC 199

RESULT 15
AAG54479
ID AAG54479 standard; Protein; 207 AA.
XX AAG54479;
AC AAG54479;
XX 18-OCT-2000 (first entry)
XX Zea mays protein fragment SEQ ID NO: 69469.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX Zea mays subsp. mays.
XX EPI033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
```

PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.



```
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      50.4%; Score 535; DB 21; Length 207;
Best Local Similarity 51.5%; Pred. No. 2.7e-49;
Matches 102; Conservative 35; Mismatches 57; Indels 4; Gaps 2;

QY 7 LFKVILLGSGYKSLMNRYYVTKFDTOLFTHTIGVEFLNKDLEVDGHFVTMQIWDTAGQ 66
Db ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 8 LKVIILGDSGVGKTSLMNQYNNRFSNQYKATIGADFLTKEVQIDDLFLQIWDTAGQ 67
Db ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 67 EFRSLRTPFYRGSDCLLTFPSVDDSQSFQNLNWKKEFYIYADVKEPESPPFVILGNKI 126
Db ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 68 ERFQSLGVAFYRGADCCVLVYDVNXXSKSFEKLNWREEFLLIQASPSDPENFPFVVLGNKI 127
Db ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 127 DI---SEROVSTEEAQACRDNGDYPFETSADATNVAAPFEAEVRRVLATEDRSD-HL 182
Db ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 128 DVDGGSRTVSEKKAKAWCASKGNIPYPETSAGEFNVEAAFECTARNVAVNEPEEDMYL 187
Db ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 183 IOTDTVNLHHRKPKPSSSC 200
Db | | | | | : | | | | |
QY 188 PDTIDVGAGRQQRSSGC 205
Db | | | | | : | | | | |
```

Search completed: February 10, 2004, 04:55:52  
Job time : 77 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 04:54:36 ; Search time 27 Seconds  
(without alignments)  
314.981 Million cell updates/sec

Title: US-09-988-974-5

Perfect score: 1062  
Sequence: 1 MAGKSLFKVILLGDGVGK.....LIQDTVNLHRKPKPSSCC 201

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/aaa/5A COMB pep:\*  
2: /cgn2\_6/prodata/2/aaa/5B COMB pep:\*  
3: /cgn2\_6/prodata/2/aaa/6A COMB pep:\*  
4: /cgn2\_6/prodata/2/aaa/6B COMB pep:\*  
5: /cgn2\_6/prodata/2/aaa/PCTUS COMB pep:\*  
6: /cgn2\_6/prodata/2/aaa/backfiles1 pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1062	100.0	201	2	US-08-773-423-5 Sequence 5, Appli
2	1057	99.5	201	2	US-08-773-423-9 Sequence 9, Appli
3	538.5	50.7	202	2	US-08-531-525-24 Sequence 24, Appl
4	538.5	50.7	202	2	US-08-718-270A-24 Sequence 24, Appl
5	527	49.6	203	2	US-08-531-525-23 Sequence 23, Appl
6	527	49.6	203	2	US-08-718-270A-23 Sequence 23, Appl
7	374	35.2	214	2	US-08-531-525-52 Sequence 52, Appl
8	374	35.2	214	2	US-08-718-270A-52 Sequence 52, Appl
9	335.5	31.6	201	2	US-08-531-525-13 Sequence 13, Appl
10	335.5	31.6	201	2	US-08-718-270A-13 Sequence 13, Appl
11	333.5	31.4	202	2	US-08-531-525-14 Sequence 14, Appl
12	333.5	31.4	202	2	US-08-718-270A-14 Sequence 14, Appl
13	327	30.8	217	2	US-08-773-423-3 Sequence 3, Appli
14	325.5	30.6	203	2	US-08-766-551-8 Sequence 8, Appli
15	323	30.4	213	2	US-08-531-525-11 Sequence 11, Appl
16	323	30.4	213	2	US-08-718-270A-11 Sequence 11, Appl
17	322	30.3	201	4	US-08-916-901-3 Sequence 3, Appli
18	322	30.3	201	4	US-09-154-603-3 Sequence 3, Appli
19	322	30.3	207	2	US-08-824-873-4 Sequence 4, Appli
20	322	30.3	207	3	US-09-198-184-4 Sequence 4, Appli
21	321	30.2	201	4	US-08-916-901-8 Sequence 8, Appli
22	321	30.2	201	4	US-09-154-603-8 Sequence 8, Appli
23	320	30.1	213	2	US-08-773-423-8 Sequence 8, Appli
24	318	29.9	204	4	US-09-149-476-363 Sequence 363, App
25	317	29.8	215	2	US-08-531-525-10 Sequence 10, Appl
26	317	29.8	215	2	US-08-718-270A-10 Sequence 10, Appl
27	311	29.3	207	2	US-08-531-525-35 Sequence 35, Appl

28	311	29.3	207	2	US-08-718-270A-35	Sequence 35, Appl
29	308.5	29.0	205	2	US-08-531-525-25	Sequence 25, Appl
30	308.5	29.0	205	2	US-08-718-270A-25	Sequence 25, Appl
31	304.5	28.7	213	2	US-08-531-525-36	Sequence 36, Appl
32	304.5	28.7	213	2	US-08-718-270A-36	Sequence 36, Appl
33	300.5	28.3	190	2	US-08-824-873-3	Sequence 3, Appli
34	300.5	28.3	190	3	US-09-198-184-3	Sequence 3, Appli
35	300.5	28.3	213	2	US-08-531-525-12	Sequence 12, Appl
36	300.5	28.3	213	2	US-08-718-270A-12	Sequence 12, Appl
37	299.5	28.2	198	2	US-08-531-525-51	Sequence 51, Appl
38	299.5	28.2	198	2	US-08-718-270A-51	Sequence 51, Appl
39	296	27.9	212	4	US-09-399-913-67	Sequence 67, Appl
40	295.5	27.8	194	3	US-08-741-411-9	Sequence 9, Appli
41	293.5	27.6	191	4	US-09-075-454-3	Sequence 18, Appl
42	293	27.6	212	2	US-08-531-525-18	Sequence 18, Appl
43	289	27.6	212	2	US-08-718-270A-18	Sequence 18, Appl
44	289	27.2	210	2	US-08-531-525-16	Sequence 16, Appl
45	289	27.2	210	2	US-08-718-270A-16	Sequence 16, Appl

## ALIGNMENTS

RESULT 1  
US-08-773-423-5  
; Sequence 5, Application US/08773423  
; Patent No. 5869291  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Bandman, Olga  
; TITLE OF INVENTION: NOVEL RAB PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA: US/08/773,423  
; APPLICATION NUMBER: US/08/773,423  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0183 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Consensus  
; CLONE: Consensus  
US-08-773-423-5

Query Match 100.0%; Score 1062; DB 2; Length 201;  
Best Local Similarity 100.0%; Pred. No. 1.8e-112;

Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGKSLFKVILLGDGSGKSLMNRVYTNKFTQTLFHTIGVEFLNKDLEVDGHFVTMQI 60  
Db 1 MAGKSLFKVILLGDGSGKSLMNRVYTNKFTQTLFHTIGVEFLNKDLEVDGHFVTMQI 60  
Qy 61 WDTAGQERFSLRTPFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYYADVKEPSFPFV 120  
Db 61 WDTAGQERFSLRTPFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYYADVKEPSFPFV 120  
Qy 121 ILGNKIDISROVSTEEAQAQWCRDNGDYPYFETSADKATNVAAPFAEVRVLATEDRSD 180  
Db 121 ILGNKIDISROVSTEEAQAQWCRDNGDYPYFETSADKATNVAAPFAEVRVLATEDRSD 180  
Qy 121 ILGNKIDISROVSTEEAQAQWCRDNGDYPYFETSADKATNVAAPFAEVRVLATEDRSD 180  
Db 121 ILGNKIDISROVSTEEAQAQWCRDNGDYPYFETSADKATNVAAPFAEVRVLATEDRSD 180  
Qy 181 HLIQTDVNLHRKPKSSSCC 201  
Db 181 HLIQTDVNLHRKPKSSSCC 201

RESULT 2  
US-08-773-423-9  
; Sequence 9, Application US/08773423  
; Patent No. 5869291  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Bandman, Olga  
; TITLE OF INVENTION: NOVEL RAB PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/773,423  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0183 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 486830  
US-08-773-423-9

Query Match 99.5%; Score 1057; DB 2; Length 201;  
Best Local Similarity 99.5%; Pred. No. 6.8e-112;  
Matches 200; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGKSLFKVILLGDGSGKSLMNRVYTNKFTQTLFHTIGVEFLNKDLEVDGHFVTMQI 60

Db 1 MAGKSLFKVILLGDGSGKSLMNRVYTNKFTQTLFHTIGVEFLNKDLEVDGHFVTMQI 60  
Qy 61 WDTAGQERFSLRTPFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYYADVKEPSFPFV 120  
Db 61 WDTAGQERFSLRTPFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYYADVKEPSFPFV 120  
Qy 121 ILGNKIDISROVSTEEAQAQWCRDNGDYPYFETSADKATNVAAPFAEVRVLATEDRSD 180  
Db 121 ILGNKIDISROVSTEEAQAQWCRDNGDYPYFETSADKATNVAAPFAEVRVLATEDRSD 180  
Qy 181 HLIQTDVNLHRKPKSSSCC 201  
Db 181 HLIQTDVNLHRKPKSSSCC 201

RESULT 3  
US-08-531-525-24  
; Sequence 24, Application US/08531525  
; Patent No. 5840683  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5840683le, John P.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,525  
; FILING DATE: 21-SEP-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 37-94  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 202 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Dictyostelium discoideum  
US-08-531-525-24

Query Match 50.7%; Score 538.5; DB 2; Length 202;  
Best Local Similarity 51.2%; Pred. No. 4.6e-53;  
Matches 104; Conservative 38; Mismatches 58; Indels 3; Gaps 3;

Qy 1 MAGKSLFKVILLGDGSGKSLMNRVYTNKFTQTLFHTIGVEFLNKDLEVDGHFVTMQI 60  
Db 1 MTKKKVLLKVIILGDSGVGKTSLMNQYVNNKFNQYKATIGADFLTKELMWDVDRVVTMQI 60  
Qy 61 WDTAGQERFSLRTPFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYYADVKEPSFPFV 120

Db 61 WDTAQERFSRLTPFYRGSDCCLLTFSDSDSQFNLNNKKEFIYYADVKEPESFPFV 120  
QY 121 ILGNKIDI-SRQVSTEEAQAQWCRDNGDYPYFETSAKDATNVAAAFEEAVRVLATEDRS 179  
Db 121 VLGNKIDLENQVWSQKGAASWCQSKGNIPYFETSAKEAINVEQAFQTIARNAIKLEDGL 180  
QY 180 DHLIQTDTVNLHRKPKPS-SSCC 201  
Db 181 VFPIPTN-IQVPEPQPAKSGCC 202

## RESULT 4

US-08-718-270A-24  
; Sequence 24, Application US/08718270A  
; Patent No. 5910478  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5910478le, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptidomimetics Inhibiting  
; TITLE OF INVENTION: the Oncogenic Action of P21 Ras  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/718,270A  
; FILING DATE: 20-SEP-1996  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/531,525  
; FILING DATE: 21-SEP-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/004,091  
; FILING DATE: 21-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 78-95  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 202 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Dictyostelium discoideum  
US-08-718-270A-24

Query Match 50.7%; Score 538.5; DB 2; Length 202;  
Best Local Similarity 51.2%; Pred. No. 4.6e-53;  
Matches 104; Conservative 38; Mismatches 58; Indels 3; Gaps 3;  
QY 1 MAGKSLFKVILGDDGGVGSLSLNRYVYTKFPTQLFHTTGVFLNKDLEVDGHFVTMQI 60  
Db 1 MTKKKVLLKVIILGDSGVGKTSLSMNQYVNNKFSNQYKATIGADFLTKELMVDVDRVTMQI 60

QY 61 WDTAQERFSRLTPFYRGSDCCLLTFSDSDSQFNLNNKKEFIYYADVKEPESFPFV 120  
Db 61 WDTAQERFSRLTPFYRGSDCCLLTFSDSDSQFNLNNKKEFIYYADVKEPESFPFV 120  
QY 121 ILGNKIDI-SRQVSTEEAQAQWCRDNGDYPYFETSAKDATNVAAAFEEAVRVLATEDRS 179  
Db 121 VLGNKIDLENQVWSQKGAASWCQSKGNIPYFETSAKEAINVEQAFQTIARNAIKLEDGL 180  
QY 180 DHLIQTDTVNLHRKPKPS-SSCC 201  
Db 181 VFPIPTN-IQVPEPQPAKSGCC 202

## RESULT 5

US-08-531-525-23  
; Sequence 23, Application US/08531525  
; Patent No. 5840683  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5840683le, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
; TITLE OF INVENTION: of P21 Ras  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,525  
; FILING DATE: 21-SEP-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 37-94  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 203 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Canis familiaris  
US-08-531-525-23

Query Match 49.6%; Score 527; DB 2; Length 203;  
Best Local Similarity 51.5%; Pred. No. 9.3e-52;  
Matches 102; Conservative 30; Mismatches 64; Indels 2; Gaps 2;

QY 4 KSSLFKVILGDDGGVGSLSLNRYVYTKFPTQLFHTTGVFLNKDLEVDGHFVTMQIWD 63  
Db 5 KKVLLKVIILGDSGVGKTSLSMNQYVNNKFSNQYKATIGADFLTKELMVDVDRVTMQIWD 64  
QY 64 AQGERFSRLTPFYRGSDCCLLTFSDSDSQFNLNNKKEFIYYADVKEPESFPFVILG 123  
Db 65 AQGERFSRLTPFYRGSDCCLLTFSDSDSQFNLNNKKEFIYYADVKEPESFPFVILG 123

QY 124 NKIDISERQVSTEEAQAACWCRDNGDYPYFETSAKDATNVAAAFEEAVRRVLATEDRSDDL 183  
Db 124 NKIDLENQVATKRAQAWCYSKNNIPYFETSAKEAINVEQAFQTTIARNALKOETEVELYN 183  
QY 184 Q-TDVTNLHRKPKPSSC 200  
Db 184 EPPEPIKLDKDAKSAEC 201

## RESULT 6

US-08-718-270A-23  
; Sequence 23, Application US/08718270A  
; Patent No. 5910478  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5910478le, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptidomimetics Inhibiting  
; TITLE OF INVENTION: the Oncogenic Action of P21 Ras  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/718,270A  
; FILING DATE: 20-SEP-1996  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/531,525  
; FILING DATE: 21-SEP-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/004,091  
; FILING DATE: 21-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 78-95  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 203 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Canis familiaris  
US-08-718-270A-23

Query Match 49.6%; Score 527; DB 2; Length 203;  
Best Local Similarity 51.5%; Pred. No. 9.3e-52;  
Matches 102; Conservative 30; Mismatches 64; Indels 2; Gaps 2;  
QY 4 KSSLFKVILLGGVGKSLMRYVTKPDTQLFITIGVEFLNKDLEVDGHVFTMQIWD 63  
Db 5 KVVLLKVVILGGVGKSLMRYVTKPDTQLFITIGVEFLNKDLEVDGHVFTMQIWD 64  
QY 64 AQGERFSLRTFPYRGSDCCLLTFSVDDSQSFQNTSNKKKEFIYYADVKEPSPFVLIG 123

Db 65 AQGERFQSLGV-FYRGADCCVLVDFVTAENTFKTLDSDWRDEFLIOASPRDPENFPFVLIG 123  
QY 124 NKIDISERQVSTEEAQAACWCRDNGDYPYFETSAKDATNVAAAFEEAVRRVLATEDRSDDL 183  
Db 124 NKIDLENQVATKRAQAWCYSKNNIPYFETSAKEAINVEQAFQTTIARNALKOETEVELYN 183  
QY 184 Q-TDVTNLHRKPKPSSC 200  
Db 184 EPPEPIKLDKDAKSAEC 201

## RESULT 7

US-08-531-525-52  
; Sequence 52, Application US/08531525  
; Patent No. 5840683  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5840683le, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
; TITLE OF INVENTION: of P21 Ras  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,525  
; FILING DATE: 21-SEP-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 37-94  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 214 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
US-08-531-525-52

Query Match 35.2%; Score 374; DB 2; Length 214;  
Best Local Similarity 38.8%; Pred. No. 2.3e-34;  
Matches 80; Conservative 44; Mismatches 66; Indels 16; Gaps 6;  
QY 7 LFKVILLGGVGKSLMRYVTKPDTQLFITIGVEFLNKDLEVDGHVFTMQIWD 66  
Db 11 LFKVILLGGVGKSLMRYVTKPDTQLFITIGVEFLNKDLEVDGHVFTMQIWD 70  
QY 67 ERFRSLRTFPYRGSDCCLLTFSVDDSQSFQNTSNKKKEFIYYADVKEPSPFVLGNKI 126  
Db 71 ERYRAITSAYRGAVGALLVYDIKHLTYENVERMLKELRDHAD----SNIVIMLVGNKS 126  
QY 127 DISE-ROVSTEEAQAACWCRDNGDYPYFETSAKDATNVAAAFEEAVRRVLATEDRSDDL- 181

Db 127 DLRLRAVPTDEARAFKNG-LSFIETSDNSTVNAAFQILTEIYRIVSQKQMSRE 185  
Qy 182 ---LIQTDVTLHRKP---KPSSSCC 201  
Db 186 NDMPSPNNVVPIHVPTTEKPKVQCC 211

RESULT 8  
US-08-718-270A-52  
; Sequence 52, Application US/08718270A  
; Patent No. 5910478  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5910478le, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptidomimetics Inhibiting  
; TITLE OF INVENTION: the Oncogenic Action of P21 Ras  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/718,270A  
; FILING DATE: 20-SEP-1996  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/531,525  
; FILING DATE: 21-SEP-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/004,091  
; FILING DATE: 21-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 78-95  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 214 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORGANISM: Homo sapiens  
; ORGANISM: Homo sapiens  
US-08-718-270A-52

Query Match 35.2%; Score 374; DB 2; Length 214;  
Best Local Similarity 38.8%; Pred. No. 2.3e-34;  
Matches 80; Conservative 44; Mismatches 66; Indels 16; Gaps 6;

Qy 7 LFKVILGDGVGKSLMRYVTNKFDTLFHTIGVEFLNKDLVDGHHFVTMOWDTAGQ 66  
Db 11 LFKVILGDSGVGKSLMRYVTNKFDTLFHTIGVEFLNKDLVDGHHFVTMOWDTAGQ 70  
Qy 67 ERFRLRTPFYRGSDCCLLTFVSDDSQSQFONLSNNKKFIYVADVKEPSPFPVILGNKI 126  
Db 71 ERYRAITSAYRGAVGALLVYDIKHLTYENVERWLKELRDHAD-----SNIVIMLVGNKS 126

Qy 127 DISE-ROVSTEEACAMCRDNGDYPYEETSAKDATNVAAAFE---EAVRVLATEDRSDH- 181  
Db 127 DLRLRAVPTDEARAFKNG-LSFIETSDNSTVNAAFQILTEIYRIVSQKQMSRE 185  
Qy 182 ---LIQTDVTLHRKP---KPSSSCC 201  
Db 186 NDMPSPNNVVPIHVPTTEKPKVQCC 211

RESULT 9  
US-08-531-525-13  
; Sequence 13, Application US/08531525  
; Patent No. 5840683  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5840683le, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
; TITLE OF INVENTION: of P21 Ras  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,525  
; FILING DATE: 21-SEP-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 37-94  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Discozyme ommata  
US-08-531-525-13

Query Match 31.6%; Score 335.5; DB 2; Length 201;  
Best Local Similarity 36.9%; Pred. No. 4.9e-30;  
Matches 76; Conservative 40; Mismatches 67; Indels 23; Gaps 6;

Qy 7 LFKVILGDGVGKSLMRYVTNKFDTLFHTIGVEFLNKDLVDGHHFVTMOWDTAGQ 66  
Db 8 LFKVILGDSGVGKSLMRYVTNKFDTLFHTIGVEFLNKDLVDGHHFVTMOWDTAGQ 67  
Qy 67 ERFRLRTPFYRGSDCCLLTFVSDDSQSQFONLSNNKKFIYVADVKEPSPFPVILGNKI 126  
Db 68 ERFRTTSSYRGAGHIIIVYDVTQDSFNKQWLQEI DRVAS-----ENNVKLLVGNKC 123  
Qy 127 DISEROV---STEEACAMCRDNGDYPYEETSAKDATNVAAAFEAVRVL-----ATED 177  
Db 124 DLTKKVDVDTKE-----FADSLGIFLETSAKNAINVQAFMTMAEIKRMGPQATSG 179

QY 178 RSDHLQTDVNLHRKPKSS--SCC 201  
Db 180 GS-----EKSNNVNIQSTPVKSSGGGCC 201

## RESULT 10

US-08-718-270A-13  
; Sequence 13, Application US/08718270A  
; Patent No. 5910478  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5910478, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptidomimetics Inhibiting  
; the Oncogenic Action of P21 Ras  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/718,270A  
; FILING DATE: 20-SEP-1996  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/531,525  
; FILING DATE: 21-SEP-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/004,091  
; FILING DATE: 21-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 78-95  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Discopyge omata  
US-08-718-270A-13

Query Match 31.6%; Score 335.5; DB 2; Length 201;  
Best Local Similarity 36.9%; Pred. No. 4.9e-30;  
Matches 76; Conservative 40; Mismatches 67; Indels 23; Gaps 6;  
QY 7 LFKVILLGDSGVGKSLMRYVTNKFDTLQFHTIGVEFLNKLEVDGHHFVTWQIWDTAGQ 66  
Db 8 LFKLLIGDSGVGKSLMRYVTNKFDTLQFHTIGVEFLNKLEVDGHHFVTWQIWDTAGQ 67  
QY 67 ERFRSLRTPYRGSDCLLTFVSQSFQNLNNKKEFIYVADVKEPESFPFVILGNKI 126  
Db 68 ERFRITTSYRGAHGIIVYVDVTDQESFNNKQWLQIEDRYAS-----ENVNKLIVGNKS 123  
QY 127 DISERQVSTEEAQAQCRDNGDYPFETSADATNVAAPFEEAVRVL-----ATED 177

Db 124 DLTKTKVVDYTKB----PADSLGIPFLETSAKATNVEQAFMTMAAEIKRKGPGNTSG 179  
QY 178 RSDHLQTDVNLHRKPKSS--SCC 201  
Db 180 GS-----EKSNNVNIQSTPVKSSGGGCC 201  
RESULT 11  
US-08-531-525-14  
; Sequence 14, Application US/08531525  
; Patent No. 5840683  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5840683, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
; the Oncogenic Action of P21 Ras  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,525  
; FILING DATE: 21-SEP-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 37-94  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 202 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Lymeia stagnalis  
US-08-531-525-14

Query Match 31.4%; Score 333.5; DB 2; Length 202;  
Best Local Similarity 34.7%; Pred. No. 8.3e-30;  
Matches 70; Conservative 46; Mismatches 69; Indels 17; Gaps 4;  
QY 7 LFKVILLGDSGVGKSLMRYVTNKFDTLQFHTIGVEFLNKLEVDGHHFVTWQIWDTAGQ 66  
Db 11 LFKLLIGDSGVGKSLMRYVTNKFDTLQFHTIGVEFLNKLEVDGHHFVTWQIWDTAGQ 70  
QY 67 ERFRSLRTPYRGSDCLLTFVSQSFQNLNNKKEFIYVADVKEPESFPFVILGNKI 126  
Db 71 ERFRITTSYRGAHGIIVYVDVTDQESFNNKQWLQIEDRYAS-----ENVNKLIVGNKS 126  
QY 127 DISERQVSTEEAQAQCRDNGDYPFETSADATNVAAPF-----EAVRVLATEDS 179  
Db 127 DLTKTKVVDYTKB----PADSLGIPFLETSAKATNVEQAFMTMAAEIKRKGPGNTSG 185  
QY 180 DLIQTDVNLHRKPKSS--SCC 201





RESULT 14  
US-08-766-551-8  
; Sequence 8, Application US/08766551  
; Patent No. 5840569  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hawkins, Phillip R.  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL HUMAN GTP-BINDING PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/766,551  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0168 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 203 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Genbank  
; CLONE: 1457955  
US-08-766-551-8

Query Match 30.6%; Score 325.5; DB 2; Length 203;  
Best Local Similarity 38.9%; Pred. No. 6.8e-29;  
Matches 77; Conservative 39; Mismatches 73; Indels 9; Gaps 5;  
  
Qy 7 LFKVILLGGGVGKSLMRYVTKNEDTQTFHTIGVEFLNKLDEVDGHPVWQIWDTAGQ 66  
Db 9 LFKVILGNAGVGKTCVRRFTQGLFPFGGATIGVGFMIKTVGEINGEKVKLQIWDTAGQ 68  
  
Qy 67 ERSRLTPFYRGSDCLLTFSQSNLNNWKEFIYVADVKEPSPFFVILGNKI 126  
Db 69 ERFPSITQSYRSANALITDYTCESFRCLPFWLREIEQVASNK----VITVLVGNKI 124  
  
Qy 127 DISE-RQVSTEEAQAQCRNGDYPFETSAAKATNVAFAFEAVRRVLATEDRSDLHIQT 185  
Db 125 DLAEERVSQRAEFSQAQDMY-YLETSAKESDNVKEFLDLACR-LISEARQNTLYNN 182  
  
Qy 186 DTVNLHKKPKPS--SCC 201  
Db 183 VSSPLFEGGKSI SYLTCC 200

RESULT 15  
US-08-531-525-11

; Sequence 11, Application US/08531525  
; Patent No. 5840683  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5840683le, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,525  
; FILING DATE: 21-SEP-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 37-94  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 213 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Arabidopsis thaliana  
US-08-531-525-11

Query Match 30.4%; Score 323; DB 2; Length 213;  
Best Local Similarity 41.0%; Pred. No. 1.4e-28;  
Matches 73; Conservative 35; Mismatches 58; Indels 12; Gaps 5;  
  
Qy 1 MAGKSS-----LRFKVLGGGVGKSLMRYVTKNEDTQTFHTIGVEFLNKLDEVDGHP 55  
Db 1 MAGYADSEYDYLFLKVLIGDSGVGKSNLSRF-TKNFNLESKSTIGVEFATKTKVEKV 59  
  
Qy 56 VTQMIDTAGQERFSLRTPFYRGSDCLLTFSQSNLNNWKEFIYVADVKEPE 115  
Db 60 VKAQIWDTAGQERVAITSAYRGAVALIYDTRHATENAARWLRELGHDT----P 115  
  
Qy 116 SFPVILGNKIDISER-QVSTEEAQAQCRNGDYPFETSAAKATNVAFAFEAVRRV 172  
Db 116 NIVVMVLGNKCDLRLHVAVKTEAKAFERESLY-FMETSALDATNVNFAFVLTQI 172

Search completed: February 10, 2004, 04:59:42  
Job time : 29 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 04:56:01 ; Search time 75 Seconds  
(without alignments)  
561.145 Million cell updates/sec

Title: US-09-988-974-5  
Perfect score: 1062  
Sequence: 1 MAGKSLFKVILLGDBGVK.....LIQTIVNLHRLKPKSSCC 201

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1062	100.0	201	9	US-09-988-974-5
2	1062	100.0	209	9	US-09-988-974-5
3	1057	99.5	221	10	US-09-988-974-5
4	839	79.0	202	10	US-09-988-974-5
5	812.5	76.5	221	10	US-09-988-974-5
6	529	49.8	208	9	US-09-925-302-534
7	524.5	49.4	206	9	US-09-925-302-534
8	518.5	48.8	209	12	US-10-369-493-1738
9	495	46.6	208	12	US-10-369-493-1738
10	473.5	44.6	250	12	US-10-369-493-1738
11	439.5	41.4	267	12	US-10-369-493-1738
12	423.5	39.9	270	12	US-10-369-493-1738
13	419.5	39.5	198	9	US-09-925-302-534
14	419.5	39.5	198	10	US-09-925-302-534
15	419.5	39.5	198	10	US-09-925-302-534

16	341	32.1	211	12	US-10-369-493-1738
17	341	32.1	223	12	US-10-369-493-1738
18	336	31.6	201	12	US-10-369-493-1738
19	336	31.6	201	12	US-10-369-493-1738
20	334.5	31.5	204	15	US-10-128-714-8214
21	333	31.4	207	12	US-10-369-493-1738
22	332	31.3	207	9	US-09-925-302-534
23	332	31.3	246	9	US-09-925-302-534
24	331.5	31.2	207	12	US-10-369-493-1738
25	329.5	31.0	222	12	US-10-369-493-1738
26	327	30.8	190	10	US-09-925-302-534
27	327	30.8	213	15	US-10-369-493-1738
28	327	30.8	217	9	US-09-988-974-5
29	326	30.7	241	12	US-10-369-493-1738
30	324.5	30.6	205	12	US-10-369-493-1738
31	324	30.5	222	10	US-09-925-302-534
32	323.5	30.5	215	12	US-10-369-493-1738
33	323	30.4	217	15	US-10-369-493-1738
34	323	30.4	239	9	US-09-925-302-534
35	322	30.3	201	10	US-09-925-302-534
36	322	30.3	201	12	US-10-369-493-1738
37	322	30.3	221	12	US-10-369-493-1738
38	322	30.3	224	15	US-10-102-806-466
39	321.5	30.3	225	15	US-10-128-714-8214
40	321	30.2	201	10	US-09-925-302-534
41	320.5	30.2	222	12	US-10-369-493-1738
42	320	30.1	213	9	US-09-988-974-5
43	319.5	30.1	206	15	US-10-128-714-8214
44	319.5	30.1	206	15	US-10-128-714-8214
45	319.5	30.1	210	12	US-10-369-493-1738

#### ALIGNMENTS

RESULT 1  
US-09-988-974-5  
; Sequence 5, Application US/09988974  
; Patent No. US20020090712A1  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; Goli, Surya K.  
; Bandman, Olga  
; TITLE OF INVENTION: NOVEL RAB PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/988,974  
; FILING DATE: 19-No. US20020090712A1-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/215,887  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0183 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 201 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: Consensus  
 CLONE: Consensus  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
 US-09-988-974-5

Query Match 100.0%; Score 1062; DB 9; Length 201;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-106;  
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAGKSSLFKVLILGDDGGVGSLLMNRVYTNKFTQTLFHTIGVFLNKDLEVDGHFTVMOI 60  
 DB 1 MAGKSSLFKVLILGDDGGVGSLLMNRVYTNKFTQTLFHTIGVFLNKDLEVDGHFTVMOI 60  
 QY 61 WDTAGQERFRLRTPFFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYADVKEPESFPFV 120  
 DB 61 WDTAGQERFRLRTPFFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYADVKEPESFPFV 120  
 QY 121 ILGNKIDISERQVSTEEAQAQWCRDNGDYPYFETSAKDNTVAAPFEAVRRVLATEDRSD 180  
 DB 121 ILGNKIDISERQVSTEEAQAQWCRDNGDYPYFETSAKDNTVAAPFEAVRRVLATEDRSD 180  
 QY 181 HLIQTDVNLHRKPKSSSCC 201  
 DB 181 HLIQTDVNLHRKPKSSSCC 201

RESULT 2  
 US-09-864-761-42996  
 ; Sequence 42996, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; FILE REFERENCE: Aecomica-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263.6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00670  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: US 60/234,687  
 PRIOR FILING DATE: 2000-09-21  
 PRIOR APPLICATION NUMBER: US 09/608,408  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: US 09/774,203  
 PRIOR FILING DATE: 2001-01-29  
 NUMBER OF SEQ ID NOS: 49117  
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
 SEQ ID NO 42996  
 LENGTH: 209  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: MAP TO AC003037.1  
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
 OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.2  
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1  
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2  
 OTHER INFORMATION: EST HUMAN HIT: BE735344.1, EVALUE 4.00e-82  
 OTHER INFORMATION: SWISSPROT HIT: P51151, EVALUE 1.00e-108  
 US-09-864-761-42996

Query Match 100.0%; Score 1062; DB 9; Length 209;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-106;  
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAGKSSLFKVLILGDDGGVGSLLMNRVYTNKFTQTLFHTIGVFLNKDLEVDGHFTVMOI 60  
 DB 9 MAGKSSLFKVLILGDDGGVGSLLMNRVYTNKFTQTLFHTIGVFLNKDLEVDGHFTVMOI 68  
 QY 61 WDTAGQERFRLRTPFFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYADVKEPESFPFV 120  
 DB 69 WDTAGQERFRLRTPFFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYADVKEPESFPFV 128  
 QY 121 ILGNKIDISERQVSTEEAQAQWCRDNGDYPYFETSAKDNTVAAPFEAVRRVLATEDRSD 180  
 DB 129 ILGNKIDISERQVSTEEAQAQWCRDNGDYPYFETSAKDNTVAAPFEAVRRVLATEDRSD 188  
 QY 181 HLIQTDVNLHRKPKSSSCC 201  
 DB 189 HLIQTDVNLHRKPKSSSCC 209

RESULT 3  
 US-09-988-974-9  
 ; Sequence 9, Application US/09889974  
 ; Patent No. US20020090712A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hillman, Jennifer L.  
 ; Goli, Surya K.  
 ; Bandman, Olga  
 ; TITLE OF INVENTION: NOVEL RAB PROTEINS  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/988,974  
 ; FILING DATE: 19-No. US20020090712A1-2001  
 ; CLASSIFICATION: <Unknown>

```
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/215,887
/ FILING DATE: <Unknown>
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Billings, Lucy J.
/ REGISTRATION NUMBER: 36,749
/ REFERENCE/DOCKET NUMBER: PF-0183 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-855-0555
/ TELEFAX: 415-845-4166
/ TELEX: <Unknown>
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 201 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: GenBank
/ CLONE: 486830
/ SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-988-974-9

Query Match 99.5%; Score 1057; DB 9; Length 201;
Best Local Similarity 99.5%; Pred. No. 3.9e-106;
Matches 200; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGKSSLFKVLILGDDGGVKGSSLMNRYVYTNKFDLTQFHTTIGVFLNKLQLEVDGHHFTVMTQI 60
Db 1 MAGKSSLFKVLILGDDGGVKGSSLMNRYVYTNKFDLTQFHTTIGVFLNKLQLEVDGHHFTVMTQI 60

Qy 61 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 120
Db 61 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 120

Qy 121 ILGNKIDISERQVSTEEAQAACWCRDNGDYPYFETSAKDATNVAAAFEEAVRRVLATEDRSD 180
Db 121 ILGNKIDISERQVSTEEAQAACWCRDNGDYPYFETSAKDATNVAAAFEEAVRRVLATEDRSD 180

Qy 181 HLIQTDTVNLHKKPKPSSSCC 201
Db 181 HLIQTDTVNLHKKPKPSSSCC 201

RESULT 4
US-09-764-868-1121
; Sequence 1121, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1121
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-1121

Query Match 79.0%; Score 839; DB 10; Length 222;
Best Local Similarity 76.1%; Pred. No. 1.8e-82;
Matches 153; Conservative 22; Mismatches 26; Indels 0; Gaps 0;

Qy 1 MAGKSSLFKVLILGDDGGVKGSSLMNRYVYTNKFDLTQFHTTIGVFLNKLQLEVDGHHFTVMTQI 60
Db 22 MSGKSLLLKVLILGDDGGVKGSSLMNRYVYTNKFDLTQFHTTIGVFLNKLQLEVDGHHFTVMTQI 81

Qy 61 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 120
Db 61 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 120

RESULT 5
US-09-764-868-702
; Sequence 702, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 702
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (221)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-702

Query Match 76.5%; Score 812.5; DB 10; Length 221;
Best Local Similarity 75.5%; Pred. No. 1.3e-79;
Matches 151; Conservative 22; Mismatches 26; Indels 1; Gaps 1;

Qy 1 MAGKSSLFKVLILGDDGGVKGSSLMNRYVYTNKFDLTQFHTTIGVFLNKLQLEVDGHHFTVMTQI 60
Db 22 MSGKSLLLKVLILGDDGGVKGSSLMNRYVYTNKFDLTQFHTTIGVFLNKLQLEVDGHHFTVMTQI 81

Qy 61 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 120
Db 61 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 120

Qy 82 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 141
Db 82 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 141

Qy 121 ILGNKIDISERQVSTEEAQAACWCRDNGDYPYFETSAKDATNVAAAFEEAVRRVLATEDRSD 180
Db 121 ILGNKIDISERQVSTEEAQAACWCRDNGDYPYFETSAKDATNVAAAFEEAVRRVLATEDRSD 180

Qy 142 VLGNKVDKEDRQVTTTEAQTWCMEGDPYLETSAKDTNVTVAFEAVRQVLAVEEQLE 201
Db 142 VLGNKVDKEDRQVTTTEAQTWCMEGDPYLETSAKDTNVTVAFEAVRQVLAVEEQLE 201

Qy 181 HLIQTDTVNLHKKPKPSSSCC 200
Db 181 HLIQTDTVNLHKKPKPSSSCC 200

Db 202 HCMLGHTIDL-TGIQAGSSC 220

RESULT 6
US-09-925-302-629
; Sequence 629, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCI/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
; LENGTH: 208
```

```
Db 82 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 141
Qy 121 ILGNKIDISERQVSTEEAQAACWCRDNGDYPYFETSAKDATNVAAAFEEAVRRVLATEDRSD 180
Db 142 VLGNKVDKEDRQVTTTEAQTWCMEGDPYLETSAKDTNVTVAFEAVRQVLAVEEQLE 201
Qy 181 HLIQTDTVNLHKKPKPSSSCC 201
Db 202 HCMLGHTIDL-TGIQAGSSC 222
```

```
RESULT 5
US-09-764-868-702
; Sequence 702, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 702
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (221)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-702
```

```
Query Match 76.5%; Score 812.5; DB 10; Length 221;
Best Local Similarity 75.5%; Pred. No. 1.3e-79;
Matches 151; Conservative 22; Mismatches 26; Indels 1; Gaps 1;

Qy 1 MAGKSSLFKVLILGDDGGVKGSSLMNRYVYTNKFDLTQFHTTIGVFLNKLQLEVDGHHFTVMTQI 60
Db 22 MSGKSLLLKVLILGDDGGVKGSSLMNRYVYTNKFDLTQFHTTIGVFLNKLQLEVDGHHFTVMTQI 81

Qy 61 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 120
Db 61 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 120

Qy 82 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 141
Db 82 WDTAGQERFRLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFPV 141

Qy 121 ILGNKIDISERQVSTEEAQAACWCRDNGDYPYFETSAKDATNVAAAFEEAVRRVLATEDRSD 180
Db 121 ILGNKIDISERQVSTEEAQAACWCRDNGDYPYFETSAKDATNVAAAFEEAVRRVLATEDRSD 180

Qy 142 VLGNKVDKEDRQVTTTEAQTWCMEGDPYLETSAKDTNVTVAFEAVRQVLAVEEQLE 201
Db 142 VLGNKVDKEDRQVTTTEAQTWCMEGDPYLETSAKDTNVTVAFEAVRQVLAVEEQLE 201

Qy 181 HLIQTDTVNLHKKPKPSSSCC 200
Db 181 HLIQTDTVNLHKKPKPSSSCC 200

Db 202 HCMLGHTIDL-TGIQAGSSC 220
```

```
RESULT 6
US-09-925-302-629
; Sequence 629, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCI/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
; LENGTH: 208
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-302-629

Query Match      49.8%; Score 529; DB 9; Length 208;
Best Local Similarity 56.6%; Pred. No. 5.8e-49;
Matches 98; Conservative 24; Mismatches 51; Indels 0; Gaps 0;

QY 4 KSSLFKVILLGDGGVGSLSMNRVYTNKFDTLFHTIGVEFLNKDLEVDGHFVTM 63
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 6 KCVLLKVIILGDSGVGKTSLSNQVYVNFKNQYKATIGADFLTKVNVDDRLVTM 65
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 64 AGQERFRSLRTPFYRGSDCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFV 123
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 66 AGQERFQSLGVAFYRGADCCVLVDVTPATNFTKTLDSWRDEFLIQASPRDPFV 125
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 124 NKIDISEROVSTEEAQAACRNGDYPYFETSAKDATNVAAAFEEAVRVLATE 176
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 126 NKIDLENROVATKQAQWCYSKNIPYFETSAKEAINVEQAFOTIARNALKQE 178
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
US-09-828-310-13
; Sequence 13, Application US/09828310
; Patent No. US200200661241
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNETT, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; TITLE OF INVENTION: GTP BINDING STRESS-RELATED PROTEINS AND METHODS OF USE
; FILE REFERENCE: 16313-0039
; CURRENT APPLICATION NUMBER: US/09/828,310
; PRIOR FILING DATE: 2001-04-06
; PRIOR FILING DATE: 60/196,001
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Physcomitrella patens
US-09-828-310-13

Query Match      49.4%; Score 524.5; DB 9; Length 206;
Best Local Similarity 51.2%; Pred. No. 1.7e-48;
Matches 104; Conservative 33; Mismatches 61; Indels 5; Gaps 3;

QY 2 AGKSLFKVILLGDGGVGSLSMNRVYTNKFDTLFHTIGVEFLNKDLEVDGHFVTM 61
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 ARKRTLLKVIILGDSGVGKTSLSNQVYVNFKNQYKATIGADFLTKVQVEDRLVTM 62
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 62 DTAGQERFRSLRTPFYRGSDCLLTFSVDDSQSFQNLNWKKEFIYYADVKEPSPFV 121
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 63 DTAGQERFQSLGVAFYRGADCCVLVDVNVNMFKFDNLNWRDEFLIQASPSQENPFV 122
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 122 LGNKIDIT---SERQVSTEEAQAACRNGDYPYFETSAKDATNVAAAFEEAVRVLATE 178
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 123 LGNKVDVGDGNSRVSEKAKANCAKAGGIPYFETSAKEDFNVDAAFOCIANKALKNETE 182
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 179 SD-HLIQTDTVNLHRRKPSSSC 200
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 183 EBIYLPDTIDVNASR-PQKTSGC 204
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 8
US-10-369-493-5598
; Sequence 5598, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Cao, Yongwei
```

QY 123 GNKIDISERO--VSTEEAQAACRDNGDYPYFETSAKDNTVAAPFEAVRVLATERSD 180  
Db 125 GNKIDAESKIVSKSQAELAKSLGDIPLFITSAXNAINWDTAFETARSALQOQADT 184  
QY 181 HLIQ---TDTVNLHRKPKPSSSCC 201  
Db 185 EAFEDDYNDAINIRLDGNNSCSC 208

RESULT 10  
US-10-369-493-2572  
; Sequence 2572, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 2572  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Schizosaccharomyces pombe  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(250)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-2572

Query Match 44.6%; Score 473.5; DB 12; Length 250;  
Best Local Similarity 39.2%; Pred. No. 7.6e-43;  
Matches 98; Conservative 39; Mismatches 64; Indels 49; Gaps 4;

QY 1 MAGKSS-LFKVILGDDGVGKSSLMNR-----YVTKKPDQ 35  
Db 1 MAGKKHLLKVIILGESVGKTSIMNQVFXNMLMLSLCKLXKLLKSVYVNRKESKD 60  
QY 36 LPHIGVEFLNKDLEVDGHFTMQIWDTAGQERFRSLRTPPYRGSDCCLLTFSVDDSQSF 95  
Db 61 YKATIGADFLTKVLDVDDKVVTLQMDTAGQERFQSLGVAFYRGADCCVLVDVNNKSF 120  
QY 96 QNLGNWKEFTYVADVKEPESFPFVILGNKIDISER-----Q 132  
Db 121 ETLDSWRDEFLIQASPNPETFPIILGNKVDVEQRMVCVXSIEFLIQSBEFLANNFR 180  
QY 133 VSTEEAQAACRDNGDYPYFETSAKDNTVAAPFEAVRVLATERSDHLIQ-TDTVNLH 191  
Db 181 FSKKALAFCAQGEIPIYFETSAKEINVQAEFETVAKALENNMDSDDIAADFTDPIHLD 240  
QY 192 RKPSPSSSCC 201  
Db 241 MESQKTSYC 250

RESULT 11  
US-10-369-493-12690  
; Sequence 12690, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 12690  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: Aspergillus nidulans  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(267)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-12690

Query Match 41.4%; Score 439.5; DB 12; Length 267;  
Best Local Similarity 36.7%; Pred. No. 4e-39;  
Matches 98; Conservative 36; Mismatches 58; Indels 75; Gaps 6;

QY 10 VILGDDGVGKSSLMNRVYT-----NKFDTLPHITIGVEFLNKD 48  
Db 1 VILGDSGVGKTSLMNQYVSYLLFPSPRAFSAHTDPTRVNKKFSGSYKATIGADFLTKE 60  
QY 49 LEVDGHFTVMTQI-----WDTAGQERFRSLRTPPYRGSDCCLLTFSVD 90  
Db 61 VLVDRLVTQARIHGTACSIHKLFWRSRWTAGQERFQSLGVAFYRGADCCVLVDVN 120  
QY 91 DQSPQNLNWKKEFTYVADVKEPESFPF-----VILGNKIDI--SERQ 132  
Db 121 NSKSFELDXWRDEFLIQASPRDPSPFVSIGPWLRSANFAFXVVIKNKIDMEESKRM 180  
QY 133 VSTEEAQAACRDNGDYPYFETSAKDNTVAAPFE-----EAVRVLAT 175  
Db 181 ISKRAMTFQSKGNIPYFETSAKEINVQAEFISPOXCYVTGTLILISAIVIARSALAQ 240

RESULT 12  
US-10-369-493-3332  
; Sequence 3332, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 3332  
; LENGTH: 270  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(270)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-3332

Query Match 39.9%; Score 423.5; DB 12; Length 270;  
Best Local Similarity 34.8%; Pred. No. 2.2e-37;



Db 61 FRALRPAYRGAQGFLVVDITSRDSFENVKKWLEELRHAD--KENVPIVLVGNKODL 118  
QY 129 SE-----ROVSTEEAQAQCRDNGDVPYFETSAKDATNVAAAPFEAVRRVLATED 177  
Db 119 EDEDELETEGQKRVVSTEEGEALAKELGALFPFNETSAKTNTNVVEAFELAREILKK-- 176  
QY 178 RSDHLIQDITVNLHR-KPKPSSCC 201  
Db 177 -----VSEVNVNLDQPAKKKKSKCC 196

Search completed: February 10, 2004, 05:01:14  
Job time : 79 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 04:46:06 ; Search time 37 seconds  
(without alignments)  
522.430 Million cell updates/sec

Title: US-09-988-974-5  
Perfect score: 1062  
Sequence: 1 MAGKSSLFKVLGGVGVK.....LIQDTVNLHKKPKSSSCC 201

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76: \*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	1062	100.0	2	G02361
2	1057	99.5	2	G36187
3	533	50.2	2	JC4107
4	529	49.8	2	B30413
5	529	49.8	2	JC5288
6	527	49.6	2	S62733
7	526.5	49.6	2	S01934
8	523.5	49.3	2	S36368
9	523	49.2	2	S33531
10	520.5	49.0	2	T12579
11	518.5	48.8	2	T26119
12	515	48.5	2	T03629
13	514	48.4	2	T40425
14	514	48.4	2	T03628
15	509	47.9	2	H96562
16	501.5	47.2	2	C96529
17	499.5	47.0	2	S39566
18	499	47.0	2	T00770
19	498	46.9	2	T03630
20	496.5	46.8	2	S39567
21	495	46.6	2	A44334
22	493	46.4	2	T04019
23	484.5	45.6	2	C84606
24	373	35.1	2	JN0056
25	373	35.1	2	E36364
26	373	35.1	2	E37169
27	369.5	34.8	2	JC4108
28	368	34.7	2	S10026
29	359.5	33.9	2	S38740

30	353	33.2	221	2	H71444	GTP-binding protei
31	352.5	33.2	203	2	S34253	GTP-binding protei
32	350	33.0	214	2	J86441	GTP-binding protei
33	349	32.9	218	2	C38625	GTP-binding protei
34	348	32.8	217	2	A86230	hypothetical prote
35	347	32.7	219	2	G84723	probable RAS type
36	346	32.6	218	2	JC2487	GTP-binding protei
37	346	32.6	218	2	A55005	GTP-binding protei
38	345	32.5	215	2	T06443	GTP-binding protei
39	344.5	32.4	217	2	S36365	GTP-binding protei
40	344	32.4	208	2	A38202	GTP-binding protei
41	343	32.3	202	2	S14430	GTP-binding protei
42	343	32.3	214	2	T14566	GTP-binding protei
43	341	32.1	211	2	T29035	hypothetical prote
44	341	32.1	223	2	S42679	GTP-binding protei
45	340.5	32.1	258	2	B86153	ARA-5 [imported] -

## ALIGNMENTS

### RESULT 1

G02361  
small GTP binding protein Rab9 - human  
C:Species: Homo sapiens (man)  
C>Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 02-Feb-2001  
C:Accession: G02361  
R:Ioannou, Y.A.; Davies, J.P.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: H01115  
A:Accession: G02361  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-201 <IOA>  
A:Cross-references: EMBL:U4103; NID:g1174146; PIDN:AACS1200.1; PID:g1174147  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; nucleotide binding; P-loop  
F:8-127/Domain: translation elongation factor Tu homology <ETU>  
F:14-21/Region: nucleotide-binding motif A (P-loop)  
F:124-127/Region: GTP-binding NKXD motif  
F:154-156/Region: GTP-binding SAK/L motif

Query Match		100.0%;	Score 1062;	DB 2;	Length 201;
Best Local Similarity		100.0%;	Pred. No. 1.6e-89;		
Matches 201;	Conservative	0;	Mismatches	0;	Indels
				0;	Gaps
QY	1	MAGKSSLFKVLGGVGVKSSLMNRYTNKFDTLQFTTIGVEFLNKDLEVDGHFVTMQI	60		
DB	1	MAGKSSLFKVLGGVGVKSSLMNRYTNKFDTLQFTTIGVEFLNKDLEVDGHFVTMQI	60		
QY	61	WTAGQERFSLRTPFYRGSDCCLLTFSDVDSQSFQNLNNKKKPIYYADYKESFPFV	120		
DB	61	WTAGQERFSLRTPFYRGSDCCLLTFSDVDSQSFQNLNNKKKPIYYADYKESFPFV	120		
QY	121	ILGNKIDISERQVSTEEAQAQWCRDNGDYPFETSADATNVAFAFEAVRVLATERDSD	180		
DB	121	ILGNKIDISERQVSTEEAQAQWCRDNGDYPFETSADATNVAFAFEAVRVLATERDSD	180		
QY	181	HLIQDTVNLHKKPKSSSCC	201		
DB	181	HLIQDTVNLHKKPKSSSCC	201		

### RESULT 2

S36187  
GTP-binding protein rab9 - dog  
C:Species: Canis lupus familiaris (dog)  
C>Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 02-Feb-2001  
C:Accession: S36187; S36364; S15605; S37290  
R:Lombardi, D.; Soldati, T.; Riederer, M.A.; Goda, Y.; Zerial, M.; Pfeiffer, S.R.  
EMBO J. 12, 677-682, 1993  
A>Title: Rab9 functions in transport between late endosomes and the trans Golgi network  
A:Reference number: S36187; MUID:93178443; PMID:8440258

A;Accession: G36187  
A;Molecule type: mRNA  
A;Residues: 1-201 <LON>  
A;Cross-references: EMBL:X56386  
R;Chavrier, P.; Vingron, M.; Sander, C.; Simons, K.; Zerial, M.  
Mol. Cell. Biol. 10, 6578-6585, 1990  
A;Title: Molecular cloning of YPT1/SEC4-related cDNAs from an epithelial cell line.  
A;Reference number: A36364; MUID:91061765; PMID:2123294  
A;Accession: C36364  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 30-201 <CHA>  
A;Cross-references: GB:X56386; NID:G2189; PIDN:CAA39797.1; PID:G2190  
C;Genetics:  
A;Gene: rab9  
C;Superfamily: ras transforming protein; translation elongation factor Tu homology  
F;8-127/Domain: GTP binding; nucleotide binding; P-loop  
F;14-21/Region: translation elongation factor Tu homology <ETU>  
F;124-127/Region: nucleotide-binding motif A (P-loop)  
F;154-156/Region: GTP-binding SAK/L motif

Query Match 99.5%; Score 1057; DB 2; Length 201;  
Best Local Similarity 99.5%; Pred. No. 4.5e-89;  
Matches 200; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGKSLFKVILLGDGSGVKSLLMNRVYTNKFTQLFHTIGVEFLNKKOLEVDGHHFTVMTQI 60  
Db 1 MAGKSLFKVILLGDGSGVKSLLMNRVYTNKFTQLFHTIGVEFLNKKOLEVDGHHFTVMTQI 60

Qy 61 WDTAGGERPRLTPPYRGSDCCLLTFSDSQFONLSNWKKEFYIYADVKEPSPFPV 120  
Db 61 WDTAGGERPRLTPPYRGSDCCLLTFSDSQFONLSNWKKEFYIYADVKEPSPFPV 120

Qy 121 ILGNKIDISERQVSTEEAQAACWRDNGDYPYFETSADKATNVAAPFEAVRRVLATEDRSD 180  
Db 121 ILGNKIDISERQVSTEEAQAACWRDNGDYPYFETSADKATNVAAPFEAVRRVLATEDRSD 180

Qy 181 HLIQTDTVNLH--RKPKPSSSCC 201  
Db 181 HLIQTDTVNLH--RKPKPSSSCC 201

RESULT 3  
JC4107  
membrane vesicle transport protein ypt C5 - Chlamydomonas reinhardtii  
N;Alternate names: ras-like yptC5 protein; small G protein yptC5  
C;Species: Chlamydomonas reinhardtii  
C;Date: 21-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 02-Feb-2001  
R;Dietmaier, W.; Fabry, S.; Huber, H.; Schmitt, R.  
Gene 158, 41-50, 1995  
A;Title: Analysis of a family of ypt genes and their products from Chlamydomonas reinhardtii  
A;Reference number: JC4105; MUID:95309723; PMID:7789809  
A;Accession: JC4107  
A;Molecule type: mRNA  
A;Residues: 1-206 <DIE>  
A;Cross-references: GB:U13170; NID:G806725; PIDN:AA82728.1; PID:G806726  
C;Comment: This protein plays an essential role in the regulation of intracellular membrane  
C;Genetics:  
A;Gene: yptC5  
A;Introns: 18/1; 60/3; 109/3; 136/3; 185/1  
C;Superfamily: ras transforming protein; translation elongation factor Tu homology  
C;Keywords: GTP binding; nucleotide binding; P-loop  
F;9-128/Domain: translation elongation factor Tu homology <ETU>  
F;15-22/Region: nucleotide-binding motif A (P-loop)  
F;37-45/Domain: effector #status predicted <EPF>  
F;125-128/Region: GTP-binding NKXD motif  
F;158-160/Region: GTP-binding SAK/L motif

Query Match 50.2%; Score 533; DB 2; Length 206;  
Best Local Similarity 51.7%; Pred. No. 3e-41;  
Matches 105; Conservative 33; Mismatches 59; Indels 5; Gaps 3;

Qy 4 KSSLFKVILLGDGSGVKSLLMNRVYTNKFTQLFHTIGVEFLNKKOLEVDGHHFTVMTQI 63  
Db 5 KRLLLKVIILGDGSGVKTSLMNQVQKFTKEYKATIGADFLTKIEIVDDKKVTTMQIWD 64

Qy 64 AGQERFSLRTPPYRGSDCCLLTFSDSQFONLSNWKKEFYIYADVKEPSPFPVILG 123  
Db 65 AGQERFSLRTPPYRGSDCCLLTFSDSQFONLSNWKKEFYIYADVKEPSPFPVILG 124

Qy 124 NKIDI---SERQVSTEEAQAACWRDNGDYPYFETSADKATNVAAPFEAVRRVLATEDRSD 180  
Db 125 NKIDENGSSRQVSEKAKAWCASKGSIPIYFETSADKATNVAAPFEAVRRVLATEDRSD 183

Qy 181 HLIQTDTVNLH--RKPKPSSSCC 201  
Db 184 ELFPDADVMNTATQKRAGCC 206

RESULT 4  
B30413  
GTP-binding protein rab7 - dog  
C;Species: Canis lupus familiaris (dog)  
C;Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 02-Feb-2001  
C;Accession: B30413; C39648  
R;Chavrier, P.; Vingron, M.; Sander, C.; Simons, K.; Zerial, M.  
Mol. Cell. Biol. 10, 6578-6585, 1990  
A;Title: Molecular cloning of YPT1/SEC4-related cDNAs from an epithelial cell line.  
A;Reference number: A36364; MUID:91061765; PMID:2123294  
A;Accession: B30413  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-207 <CHA>  
A;Cross-references: GB:M35522; NID:G164057; PIDN:AAA30890.1; PID:G164058  
R;Chavrier, P.; Parton, R.G.; Hauri, H.P.; Simons, K.; Zerial, M.  
Cell 62, 317-329, 1990  
A;Title: Localization of low molecular weight GTP binding proteins to exocytic and endo  
A;Reference number: A39648; MUID:90322428; PMID:2115402  
A;Accession: C39648  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-207 <CH2>  
A;Cross-references: GB:M35522; NID:G164057; PIDN:AAA30890.1; PID:G164058  
C;Superfamily: ras transforming protein; translation elongation factor Tu homology  
C;Keywords: GTP binding; membrane trafficking; nucleotide binding; P-loop  
F;9-128/Domain: translation elongation factor Tu homology <ETU>  
F;15-22/Region: nucleotide-binding motif A (P-loop)  
F;125-128/Region: GTP-binding NKXD motif  
F;155-157/Region: GTP-binding SAK/L motif

Query Match 49.8%; Score 529; DB 2; Length 207;  
Best Local Similarity 56.6%; Pred. No. 7.1e-41;  
Matches 98; Conservative 24; Mismatches 51; Indels 0; Gaps 0;

Qy 4 KSSLFKVILLGDGSGVKSLLMNRVYTNKFTQLFHTIGVEFLNKKOLEVDGHHFTVMTQI 63  
Db 5 KRLLLKVIILGDGSGVKTSLMNQVQKFTKEYKATIGADFLTKIEIVDDKKVTTMQIWD 64

Qy 64 AGQERFSLRTPPYRGSDCCLLTFSDSQFONLSNWKKEFYIYADVKEPSPFPVILG 123  
Db 65 AGQERFSLRTPPYRGSDCCLLTFSDSQFONLSNWKKEFYIYADVKEPSPFPVILG 124

Qy 124 NKIDI---SERQVSTEEAQAACWRDNGDYPYFETSADKATNVAAPFEAVRRVLATEDRSD 176  
Db 125 NKIDENRQVATKRAQACWCYKKNIPYFETSADKATNVAAPFEAVRRVLATEDRSD 177

RESULT 5  
JC5268  
GTPase Rab7 protein - human  
C;Species: Homo sapiens (man)  
C;Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 02-Feb-2001  
C;Accession: JC5268  
R;Vitelli, R.; Chiariello, M.; Lattero, D.; Bruni, C.B.; Bucci, C.

Biochem. Biophys. Res. Commun. 229, 887-890, 1996  
A/Title: Molecular cloning and expression analysis of the human Rab7 GTP-ase complemented  
A/Reference number: JC5268; MUID:97115674; PMID:8954989  
A/Accession: JC5268  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-207 <VIT>  
A/Cross-references: EMBL:X93499; NID:g1089892; PIDN:CAA63763.1; PID:g1089893  
C/Suprafamily: ras transforming protein; translation elongation factor Tu homology  
C/Keywords: GTP binding; nucleotide binding; P-loop  
F:9-128/Domain: translation elongation factor Tu homology <ETU>  
F:15-22/Region: nucleotide-binding motif A (P-loop)  
F:125-128/Region: GTP-binding NKXD motif  
F:155-157/Region: GTP-binding SAK/L motif

Query Match 49.8%; Score 529; DB 2; Length 207;  
Best Local Similarity 56.6%; Pred. No. 7.1e-41;  
Matches 98; Conservative 24; Mismatches 51; Indels 0; Gaps 0;

Qy 4 KSSLFKVILLGDGGVGSLSMRYVTKFTQFHTIGVFLNKDLEVDGHFVTMQIWD 63  
Db 5 KKVLLKVIILGDSGVGKTSLSMRYVTKFTQFHTIGVFLNKDLEVDGHFVTMQIWD 64

Qy 64 AGQERFSLRTPFYRGSDCCLLTFSDSOFNLNWKKEFYIYADVKEPSPFPVILG 123  
Db 65 AGQERFSLGVAFYRGADCCVLVDVTPNTFTKLSWRDEFLIQSPDPENFPFVILG 124

Qy 124 NKIDISQVSTEAQAWCRDNGDYPFETSAKDNTVAFAAEVRRVLATE 176  
Db 125 NKIDLENQVATKRAQAWCYKNNIPYFETSAKEINVEQAFQTIARNALKQE 177

RESULT 6  
S62733  
small GTP-binding protein Rab7 - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 02-Feb-2001  
C/Accession: S62733  
R/Vitelli, R.; Chiarlotti, M.; Bruni, C.B.; Bucci, C.  
Biochim. Biophys. Acta 1264, 268-270, 1995  
A/Title: Cloning and expression analysis of the murine Rab7 cDNA.  
A/Reference number: S62733; MUID:96138545; PMID:8547311  
A/Accession: S62733  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-207 <VIT>  
A/Cross-references: EMBL:X89650; NID:g1050550; PIDN:CAA61797.1; PID:g1050551  
C/Genetics:  
A/Genes: Rab7  
C/Suprafamily: ras transforming protein; translation elongation factor Tu homology  
C/Keywords: GTP binding; nucleotide binding; P-loop  
F:9-128/Domain: translation elongation factor Tu homology <ETU>  
F:15-22/Region: nucleotide-binding motif A (P-loop)  
F:125-128/Region: GTP-binding NKXD motif  
F:155-157/Region: GTP-binding SAK/L motif

Query Match 49.6%; Score 527; DB 2; Length 207;  
Best Local Similarity 56.6%; Pred. No. 1.1e-40;  
Matches 98; Conservative 24; Mismatches 51; Indels 0; Gaps 0;

Qy 4 KSSLFKVILLGDGGVGSLSMRYVTKFTQFHTIGVFLNKDLEVDGHFVTMQIWD 63  
Db 5 KKVLLKVIILGDSGVGKTSLSMRYVTKFTQFHTIGVFLNKDLEVDGHFVTMQIWD 64

Qy 64 AGQERFSLRTPFYRGSDCCLLTFSDSOFNLNWKKEFYIYADVKEPSPFPVILG 123  
Db 65 AGQERFSLGVAFYRGADCCVLVDVTPNTFTKLSWRDEFLIQSPDPENFPFVILG 124

Qy 124 NKIDISQVSTEAQAWCRDNGDYPFETSAKDNTVAFAAEVRRVLATE 176  
Db 125 NKIDLENQVATKRAQAWCYKNNIPYFETSAKEINVEQAFQTIARNALKQE 177

RESULT 7  
S01934  
GTP-binding protein, 23K - rat  
N/Alternate names: ras-related protein, 23K  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 02-Feb-2001  
C/Accession: S01934  
R/Bucci, C.; Frunzio, R.; Chiarlotti, L.; Brown, A.L.; Rechler, M.M.; Bruni, C.B.  
Nucleic Acids Res. 16, 9979-9993, 1988  
A/Title: A new member of the ras gene superfamily identified in a rat liver cell line.  
A/Reference number: S01934; MUID:89057527; PMID:3057452  
A/Accession: S01934  
A/Molecule type: mRNA  
A/Residues: 1-201 <BUG>  
A/Cross-references: EMBL:X12535  
C/Suprafamily: ras transforming protein; translation elongation factor Tu homology  
C/Keywords: GTP binding; membrane trafficking; nucleotide binding; P-loop  
F:9-122/Domain: translation elongation factor Tu homology <ETU>  
F:9-16/Region: nucleotide-binding motif A (P-loop)  
F:119-122/Region: GTP-binding NKXD motif  
F:149-151/Region: GTP-binding SAK/L motif

Query Match 49.6%; Score 526.5; DB 2; Length 201;  
Best Local Similarity 50.5%; Pred. No. 1.2e-40;  
Matches 100; Conservative 32; Mismatches 63; Indels 3; Gaps 1;

Qy 7 LFKVILLGDGGVGSLSMRYVTKFTQFHTIGVFLNKDLEVDGHFVTMQIWDTAGQ 66  
Db 2 LLKVIILGDSGVGKTSLSMRYVTKFTQFHTIGVFLNKDLEVDGHFVTMQIWDTAGQ 61

Qy 67 ERFSLRTPFYRGSDCCLLTFSDSOFNLNWKKEFYIYADVKEPSPFPVILG 126  
Db 62 ERFSLRTPFYRGSDCCLLTFSDSOFNLNWKKEFYIYADVKEPSPFPVILG 121

Qy 127 DISERQVSTEAQAWCRDNGDYPFETSAKDNTVAFAAEVRRVLATE 183  
Db 122 DLENQVATKRAQAWCYKNNIPYFETSAKEINVEQAFQTIARNALKQEVEVLYNEFP 181

Qy 184 QDVTNHLRKPSPSSCC 201  
Db 182 EPIKLDKNERAKAESC 199

RESULT 8  
S36368  
GTP-binding protein yptV5 - Volvox carteri  
C/Species: Volvox carteri  
C/Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 02-Feb-2001  
C/Accession: S36368  
R/Fabry, S.; Jacobsen, A.; Huber, H.; Palme, K.; Schmitt, R.  
Curr. Genet. 24, 229-240, 1993  
A/Title: Structure, expression, and phylogenetic relationships of a family of ypt genes  
A/Reference number: S36368; MUID:94037148; PMID:8221932  
A/Accession: S36368  
A/Molecule type: DNA  
A/Residues: 1-205 <FAB>  
A/Cross-references: GB:L08131; NID:g409167; PIDN:AAA34254.1; PID:g409168  
C/Genetics:  
A/Genes: yptV5  
A/Introns: 18/2; 60/3; 109/3; 135/3; 184/2  
C/Suprafamily: ras transforming protein; translation elongation factor Tu homology  
C/Keywords: GTP binding; nucleotide binding; P-loop  
F:9-128/Domain: translation elongation factor Tu homology <ETU>  
F:15-22/Region: nucleotide-binding motif A (P-loop)  
F:61-67/Region: nucleotide-binding motif A (P-loop)  
F:122-129/Region: GTP binding #status predicted  
F:150-156/Region: GTP binding #status predicted

Query Match 49.3%; Score 523.5; DB 2; Length 205;  
Best Local Similarity 50.5%; Pred. No. 2.2e-40;  
Matches 102; Conservative 34; Mismatches 61; Indels 5; Gaps 3;

Qy 4 KSSLFKVILLGDGGVGSLSMRYVTKFTQFHTIGVFLNKDLEVDGHFVTMQIWD 63

Db 5 KRVLLKIIILGDSGVGKTSLSNQVYQKFTKEYKATIGADPLTKIEIVDDKKVTMQLNDT 64  
QY 64 AGQERFSLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYADVKEPESFPFVILG 123  
Db 65 AGQERFQSLGAFYRGADCCMLVFDVNNAKSFDDLNDNRDEFITQAGFSDPFDNPFVILG 124  
QY 124 NKIDI---SERQVSTEEAQAACRDNGDYPYFETSADKATNVAAPFEAVRRVLATEDRSDH 181  
Db 125 NKIDVGNVRQVTEKAKAWCASKGSIYFETSADKEDINVEAFTCTRNALRNE-KEEE 183  
QY 182 LIQDTVNLHRK--PKPSSSCC 201  
Db 184 LFVPDAVDMTSATQKRGGCC 205

## RESULT 9

S33531  
GTP-binding protein rab - garden pea  
C:Species: Pisum sativum (garden pea)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 02-Feb-2001  
C:Accession: S33531; S25543  
R:Drew, J.B.; Bown, D.; Gatehouse, J.A.  
Plant Mol. Biol. 21, 1195-1199, 1993  
A:Title: Sequence of a novel plant ras-related cDNA from Pisum sativum.  
A:Reference number: S33531; MUID:93257636; PMID:8490139  
A:Accession: S33531  
A:Molecule type: mRNA  
A:Residues: 1-206 <DRE>  
A:Cross-references: EMBL:X65650; NID:q20755; PIDN:CAA46600.1; PID:q20756  
A:Note: the authors translated the codon AAA for residue 48 as Thr, and CCC for residue 49  
C:Genetics:  
A:Gene: rab  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; nucleotide binding; P-loop  
F:9-128/Domain: translation elongation factor Tu homology <ETU>  
F:15-22/Region: nucleotide-binding motif A (P-loop)  
F:125-128/Region: GTP-binding NKXD motif  
F:158-160/Region: GTP-binding SAK/L motif

Query Match 49.2%; Score 523; DB 2; Length 206;  
Best Local Similarity 48.3%; Pred. No. 2.5e-40;  
Matches 97; Conservative 42; Mismatches 58; Indels 4; Gaps 2;  
QY 4 KSSLFKVILLGDSGVGKSSLMNRYVYTNKFDTLQFHTIGVEFLNKLQLEVDGHFVTMQLNDT 63  
Db 5 RRTLLKIIILGDSGVGKTSLSNQVYQKFTKEYKATIGADPLTKIEIVDDKKVTMQLNDT 64  
QY 64 AGQERFSLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYADVKEPESFPFVILG 123  
Db 65 AGQERFQSLGAFYRGADCCVLYVDVNSVKSFNLDNNWREEFLLIQANPSDPENPFVILG 124  
QY 124 NKIDI---SERQVSTEEAQAACRDNGDYPYFETSADKATNVAAPFEAVRRVLATEDRSD 180  
Db 125 NKIDVGNVRQVTEKAKAWCASKGSIYFETSADKEDINVEAFTCTRNALRNE-KEE 183  
QY 181 HLIQDTVNLHRKPKPSSSCC 201  
Db 184 ELYLPDITDVGNSSQPRSTGC 204

## RESULT 10

T12579  
GTP-binding protein Rab7a - common ice plant  
C:Species: Mesembryanthemum crystallinum (common ice plant)  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 21-Jan-2000  
C:Accession: T12579  
R:Michalowski C.B.; Bohnert, H.J.  
Submitted to the EMBL Data Library, January 1997  
A:Description: Sequence of a GTP-binding protein from Mesembryanthemum crystallinum.  
A:Reference number: Z17539  
A:Accession: T12579  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 1-207 <MIC>  
A:Cross-references: EMBL:U87142; NID:gi842068; PID:gi842069  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; P-loop  
F:9-128/Domain: translation elongation factor Tu homology <ETU>

Query Match 49.0%; Score 520.5; DB 2; Length 207;  
Best Local Similarity 48.1%; Pred. No. 4.2e-40;  
Matches 98; Conservative 41; Mismatches 56; Indels 5; Gaps 3;

QY 4 KSSLFKVILLGDSGVGKSSLMNRYVYTNKFDTLQFHTIGVEFLNKLQLEVDGHFVTMQLNDT 63  
Db 5 RRLKIIILGDSGVGKTSLSNQVYQKFTKEYKATIGADPLTKIEIVDDKKVTMQLNDT 64  
QY 64 AGQERFSLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYADVKEPESFPFVILG 123  
Db 65 AGQERFQSLGAFYRGADCCVLYVDVNSVKSFNLDNNWREEFLLIQANPSDPENPFVILG 124  
QY 124 NKIDI---SERQVSTEEAQAACRDNGDYPYFETSADKATNVAAPFEAVRRVLATEDRSD 180  
Db 125 NKIDVGGGRVSEKAKAWCASKGSIYFETSADKEDINVEAFTCTRNALRNE-KEE 184  
QY 181 HLIQDTVNLHRKPKP-SSSCC 201  
Db 185 TYL-EDTIDMAGSTRPQSSSAC 205

## RESULT 11

T26119  
hypothetical protein W03C9.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T26119  
R:Gregory, J.; Ainscough, R.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: Z20155  
A:Accession: T26119  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-209 <WIL>  
A:Cross-references: EMBL:Z66516; PIDN:CAA91357.1; GSPDB:GNO0020; CESP:W03C9.3  
A:Experimental source: clone W03C9  
C:Genetics:  
A:Gene: CESP:W03C9.3  
A:Map position: 2  
A:Introns: 39/3; 61/3; 112/2; 181/3  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology

Query Match 48.8%; Score 518.5; DB 2; Length 209;  
Best Local Similarity 48.8%; Pred. No. 6.5e-40;  
Matches 101; Conservative 43; Mismatches 56; Indels 7; Gaps 4;

QY 1 MAG--KSLFKVILLGDSGVGKSSLMNRYVYTNKFDTLQFHTIGVEFLNKLQLEVDGHFVTM 58  
Db 1 MSGTRKALLKVIILGDSGVGKTSLSNQVYQKFTKEYKATIGADPLTKIEIVDDKKVTM 60  
QY 59 QIWTAGQERFSLRTPFYRGSDCCLLTFSVDDSQSFQNLNWKKEFIYADVKEPESFP 118  
Db 61 QIWTAGQERFQSLGAFYRGADCCVLYVDVNSVKSFNLDNNWREEFLLIQANPSDPEN 120  
QY 119 FVILGNKIDI--SERQVSTEEAQAACRDNGDYPYFETSADKATNVAAPFEAVRRVLATE 176  
Db 121 FVLLGNKVLDLESSQRAVSSKRAQSWCQTKGNIPYVEVSAKEALNVEAFLAIRDALARES 180  
QY 177 -DRSDHLIQTDTVNL--HRKPKPSSSC 200  
Db 181 QETNDPFEPFDQIRLNFNQNNQNSGC 207

## RESULT 12

T03629  
GTP-binding protein Rab7b - common tobacco

C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 02-Feb-2001  
C:Accession: T03629  
R:Haizel, T.; Merkle, T.; Turck, F.; Nagy, F.  
Plant Physiol. 108, 59-67, 1995  
A:Title: Characterization of membrane-bound small GTP-binding proteins from Nicotiana ta  
A:Reference number: Z14896; MUID:95303981; PMID:7784525  
A:Accession: T03629  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-205 <HA1>  
A:Cross-references: EMBL:L29275; NID:G623589; PIDN:AAA74119.1; PID:G623590  
A:Experimental source: strain SRI  
C:Genetics:  
A:Gene: Rab7b  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
F:8-127/Domain: translation elongation factor Tu homology <ETU>  
F:14-21/Region: nucleotide-binding motif A (P-loop)  
F:124-127/Region: nucleotide-binding motif A (P-loop)  
F:157-159/Region: GTP-binding SAX/L motif

Query Match 48.5%; Score 515; DB 2; Length 205;  
Best Local Similarity 46.6%; Pred. No. 1.3e-39;  
Matches 95; Conservative 44; Mismatches 61; Indels 4; Gaps 2;

QY 1 MAGKSLFKVILLGDGVGKSSLMNRYVTKFDTQLFHTIGVEFLNKDLEVDGHVFTMQI 60  
DB 1 MFSANLVKVIILGDSGVGKTSIMNQYVNRKFSQYKATIGADFLTKVEQVDFDLTLQI 60

QY 61 WDTAGQERFSLRTPFYRGSDCLLTFSVDSQSFLNWNKKEFYIYADVKEPSPFV 120  
DB 61 WDTAGQERFQSLGVAFYRGADCCVLYVDVNMKSFENLNWREFFLIQASPSDPENPFV 120

QY 121 ILGNKIDI---SERQVSTEEAQWCRDNGDYFYFETSAKDATNVAAAFEEAVRRVLATED 177  
DB 121 VLGNKVIDDGNRVRVSEKKARAWCASKGNIPYFETSAKEGTFVVEAFQCIANKLKSGE 180

QY 178 RSDHLIQDTVNLHRKPKPSSSCC 201  
DB 181 -EEIYLPDTLIDVGTSGQPTGGC 203

RESULT 13  
T40425  
ras-related protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Jan-2000  
C:Accession: T40425  
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, March 1999  
A:Reference number: Z21928  
A:Accession: T40425  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-205 <SBE>  
A:Cross-references: EMBL:AL035655; PIDN:CAB38603.1; GSPDB:GN00067; SPDB:SPBC405.04C  
A:Experimental source: strain 972h-; cosmid c405  
C:Genetics:  
A:Gene: SPDB:SPBC405.04C  
A:Map position: 2  
A:Insertions: 27/3; 135/3  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology

Query Match 48.4%; Score 514; DB 2; Length 205;  
Best Local Similarity 48.8%; Pred. No. 1.6e-39;  
Matches 100; Conservative 37; Mismatches 64; Indels 4; Gaps 3;

QY 1 MAGKSS-LFKVILLGDGVGKSSLMNRYVTKFDTQLFHTIGVEFLNKDLEVDGHVFTMQ 59  
DB 1 MAGKKHLLKVIILGESGVGKTSIMNQYVNRKFSQYKATIGADFLTKVEVLDKVVTLQ 60

QY 60 IWDTAGQERFSLRTPFYRGSDCLLTFSVDSQSFLNWNKKEFYIYADVKEPSPFV 119

DB 61 LWDTAGQERFQSLGVAFYRGADCCVLYVDVNMKSFETLDSWRDEFLIQASPSNPETFPF 120

QY 120 VILGNKIDISE--RQVSTEEAQWCRDNGDYFYFETSAKDATNVAAAFEEAVRRVLATED 177  
DB 121 ILGNKVIDVEEQKRVVSKSKALAFQARGEIYFYFETSAKEAINVQAFETVAKALENWD 180

QY 178 RSDHLIQ-TDVTNLHRKPKPSSSCC 201  
DB 181 SDDIAADFTDPIHLDMESQRTSCYC 205

RESULT 14  
T03628  
GTP-binding protein Rab7a - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 02-Feb-2001  
C:Accession: T03628  
R:Haizel, T.; Merkle, T.; Turck, F.; Nagy, F.  
Plant Physiol. 108, 59-67, 1995  
A:Title: Characterization of membrane-bound small GTP-binding proteins from Nicotiana t  
A:Reference number: Z14896; MUID:95303981; PMID:7784525  
A:Accession: T03628  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-206 <HA1>  
A:Cross-references: EMBL:L29274; NID:G623587; PIDN:AAA74118.1; PID:G623588  
A:Experimental source: strain SRI  
C:Genetics:  
A:Gene: Rab7a  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
F:9-128/Domain: translation elongation factor Tu homology <ETU>  
F:15-22/Region: nucleotide-binding motif A (P-loop)  
F:125-128/Region: GTP-binding NKXD motif  
F:158-160/Region: GTP-binding SAX/L motif

Query Match 48.4%; Score 514; DB 2; Length 206;  
Best Local Similarity 47.8%; Pred. No. 1.7e-39;  
Matches 97; Conservative 40; Mismatches 62; Indels 4; Gaps 2;

QY 2 AGKSLFKVILLGDGVGKSSLMNRYVTKFDTQLFHTIGVEFLNKDLEVDGHVFTMQI 61  
DB 3 ARRRMLLKVIILGDSGVGKTSIMNQYVNRKFSQYKATIGADFLTKVEQVDFDLTLQI 62

QY 62 DTAGQERFSLRTPFYRGSDCLLTFSVDSQSFLNWNKKEFYIYADVKEPSPFV 121  
DB 63 DTAGQERFQSLGVAFYRGADCCVLYVDVNMKSFENLNWREFFLIQASPSDPENPFIV 122

QY 122 LGNKIDI---SERQVSTEEAQWCRDNGDYFYFETSAKDATNVAAAFEEAVRRVLATEDR 178  
DB 123 LGNKIDVDGNSRVVSEKKARAWCASKGNIPYFETSAKEGTFVVEAFQCIANKLKNP- 181

QY 179 SDHLIQDTVNLHRKPKPSSSCC 201  
DB 182 EDEIYLPDTIDVAGSGSQSRSTGC 204

RESULT 15  
H96562  
hypothetical protein F19K6.10 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: H96562  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali  
Rizzo, M.; Roney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: H96562  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-206 <STO>  
A/Cross-references: GB:AB005173; NID:g10645452; PIDN:AAG21568.1; GSPDB:GN00141  
C/Genetics:  
A/Gene: F19K6.10  
A/Map position: 1  
C/Superfamily: ras transforming protein; translation elongation factor Tu homology

Query Match 47.9%; Score 509; DB 2; Length 206;  
Best Local Similarity 48.5%; Pred. No. 4.7e-39;  
Matches 96; Conservative 37; Mismatches 61; Indels 4; Gaps 2;

Qy 7 LFKVILLGDGGVGVKSSLMNRYVTNKEDTQLFHTIGVEFLNKDLEVDGHFVTMQIWDTAGQ 66  
Db 8 LLKVILLGDGSGVGTSLMNGFVNKFSNQYKATIGADFLTKVEQIDDRIFTLQIWDTAGQ 67

Qy 67 ERFRSLRTPFYRGSDCCLLTFSVDDSQSFQNLGNWKKFIYYADVKPEPSFPFVILGNKI 126  
Db 68 ERFQSLGVAFYRGADCCVLVDVNVNKSFDNLNNWREEFLIQASPDPEPNFPFVILGNKT 127

Qy 127 DI---SROVSTEEAQACRDNGDYPYFETSADATNVAAFEAEVRRVLATEDRSCHLI 183  
Db 128 DVDGKSRVSEKKAACAKGKNIPYFETSAKEGPNVDAAFEICITKNAPKNEPEEPYL 187

Qy 184 QDVTNHLRHKPKPSSCC 201  
Db 188 -PDTIDVAGGQQQRSTGC 204

Search completed: February 10, 2004, 04:59:00  
Job time : 39 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 03:56:11 ; Search time 24 Seconds  
(without alignments)  
393.849 Million cell updates/sec

Title: US-09-988-974-5  
Perfect score: 1062  
Sequence: 1 MACKSLFKVLLGGGVGK.....LIQTDTVNLHRKPKSSCC 201

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1062	100.0	201	1 RB9A HUMAN	P51151 homo sapien
2	1057	99.5	201	1 RB9A CANFA	P24408 canis faml
3	1033	97.3	201	1 RB9A RAT	Q99775 rattus norv
4	839	79.0	201	1 RB9B HUMAN	Q99775 rattus norv
5	535.5	50.4	203	1 RB7 DICDI	P36411 dictyosteli
6	533	50.2	206	1 YPT5 CHLRE	Q39573 chlamydomon
7	531	50.0	206	1 RB7 PENCL	Q40787 pennisetum
8	529.5	49.9	207	1 RB7 RAT	P09527 rattus norv
9	529	49.8	207	1 RB7 CANFA	P18067 canis faml
10	529	49.8	207	1 RB7 HUMAN	P51149 homo sapien
11	527	49.6	207	1 RB7 MOUSE	P51150 mus musculu
12	524	49.3	207	1 RB7 GOSHI	Q9X988 gossypium h
13	523.5	49.3	205	1 YPT5 VOLCA	P36864 volvox cart
14	523	49.2	206	1 RB7 PEA	P31022 pisum sativ
15	520.5	49.0	207	1 RB7 MESCR	P93267 mesembryant
16	519	48.9	207	1 RB7 RABIT	Q97572 oryctolagus
17	514	48.4	205	1 YPT7 SCHPO	Q94655 schizosacch
18	509.5	48.0	205	1 RB7 NEUCR	Q9C318 neurospora
19	505	47.6	207	1 RB7 PRUAR	O24451 prunus arne
20	499	47.0	203	1 RB7 ARATH	O04157 arabidopsis
21	496	46.8	206	1 RB7 VIGAC	Q41640 vigna aconl
22	495	46.6	208	1 YPT7 YEAST	P32939 saccharomyc
23	493.5	46.5	206	1 RB7 SOYBN	Q43463 glycine max
24	373	35.1	216	1 R11A HUMAN	P24410 homo sapien
25	369.5	34.8	216	1 YPT6 CHLRE	Q39572 chlamydomon
26	368	34.7	214	1 YPT3 SCHPO	P17510 schizosacch
27	359.5	33.9	202	1 RIC1 ORYSA	P40392 oryza sativ
28	350	33.0	214	1 ARAB ARATH	P28187 arabidopsis
29	349	32.9	218	1 R11B DISOM	P22129 discopyge o
30	348	32.8	217	1 RB1C ARATH	Q04486 arabidopsis
31	346	32.6	218	1 R11B HUMAN	Q15907 homo sapien
32	346	32.6	218	1 R11B MOUSE	P46638 mus musculu
33	344.5	32.4	217	1 YPT2 VOLCA	P36861 volvox cart

34	344	32.4	208	1 YPT1 MAIZE	P16976 zea mays (m
35	343	32.3	214	1 RAB2 BETVU	Q39434 beta vulgar
36	342.5	32.3	218	1 R11E LOTJA	Q40195 lotus japon
37	341	32.1	223	1 YP31 YEAST	P38555 saccharomyc
38	340.5	32.1	258	1 ARAS ARATH	P28188 arabidopsis
39	337.5	31.8	203	1 YPT1 VOLCA	P31584 volvox cart
40	337.5	31.8	203	1 YPT2 MAIZE	Q05737 zea mays (m
41	337	31.7	207	1 R98B RAT	P70550 rattus norv
42	336.5	31.7	218	1 YPT3 NICPL	Q01111 nicotiana p
43	336	31.6	202	1 RAB1 DISOM	P22125 discopyge o
44	334.5	31.5	217	1 RIC2 ORYSA	P40393 oryza sativ
45	333.5	31.4	216	1 ARAB ARATH	P28186 arabidopsis

ALIGNMENTS

RESULT 1

RB9A\_HUMAN

ID RB9A\_HUMAN STANDARD; PRT; 201 AA.

AC P51151;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Ras-related protein Rab-9A (Rab-9).

GN RAB9A OR RAB9.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Carnivora; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92721569; PubMed=9126495;

RA Davies J.P., Cotter P.D., Ioannou Y.A.;

RT "Cloning and mapping of human Rab7 and Rab9 cDNA sequences and

RT identification of a Rab9 pseudogene.";

RL Genomics 41:131-134 (1997).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Muscle;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Marra M.A.;

RA Schnerch A., Schein J.E., Jones S.J.W., Skalska U., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

CC -1- FUNCTION: INVOLVED IN THE TRANSPORT OF PROTEINS BETWEEN

CC THE ENDOSOMES AND THE TRANS GOLGI NETWORK.

CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR EMBL; U44103; AAC51200.1; -.
DR EMBL; BC017265; AAH17265.1; -.
DR PIR; G02361; G02361.
DR HSP; P01112; IPLJ.
DR Genew; HGNC:9792; RAB9A.
DR MIM; 300284; -.
DR GO; GO:0003928; F:RAB small monomeric GTPase activity; TAS.
DR GO; GO:0006899; P:non-selective vesicle transport; TAS.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmfrng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small GTP; 1.
KW GTP-binding; Lipoprotein; Prenylation; Protein transport; Golgi stack.
FT NP_BIND 14 21 GTP (BY SIMILARITY).
FT NP_BIND 62 66 GTP (BY SIMILARITY).
FT NP_BIND 124 127 GTP (BY SIMILARITY).
FT DOMAIN 36 44 EFECTOR REGION (BY SIMILARITY).
FT DOMAIN 200 200 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 201 201 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 201 AA; 22837 MW; 65B502C21E97DB72 CRC64;

Query Match 100.0%; Score 1062; DB 1; Length 201;
Best Local Similarity 100.0%; Pred. No. 3.1e-87;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGKSLFKVILLGDGGVGKSSLMNRYVTKNFKPTQLFHTIGVEFLNKDLEVDGHFVTMQI 60
DB 1 MAGKSLFKVILLGDGGVGKSSLMNRYVTKNFKPTQLFHTIGVEFLNKDLEVDGHFVTMQI 60
QY 61 WDTAGQERFRLTPFPYRGSDCCLLTFSDVSDSOFNLNWKKEFIYYADVKEPSPFFV 120
DB 61 WDTAGQERFRLTPFPYRGSDCCLLTFSDVSDSOFNLNWKKEFIYYADVKEPSPFFV 120
QY 121 ILGNKIDISEROVSTEAQAACWCRDNGDYPYFETSAKDATNVAFAFEAVRVLATEDRSD 180
DB 121 ILGNKIDISEROVSTEAQAACWCRDNGDYPYFETSAKDATNVAFAFEAVRVLATEDRSD 180
QY 181 HLIQTDVNLHRRKPKSSCC 201
DB 181 HLIQTDVNLHRRKPKSSCC 201

RESULT 2
RB9A_CANFA STANDARD; PRT; 201 AA.
AC P24408;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ras-related protein Rab-9A (Rab-9).
GN RAB9A OR RAB9.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Cocker spaniel;
RX MEDLINE=93178443; PubMed=8440258;
RA Lombardi D., Soldati T., Riederer M.A., Goda Y., Zerial M.,
RA Pfeiffer S.R.;
RT "Rab9 functions in transport between late endosomes and the trans
RL Golgi network";
RL ENBO J. 12:677-682(1993).
RN [2]
RP SEQUENCE OF 30-201 FROM N.A.
RC STRAIN=Cocker spaniel;
RX MEDLINE=91061765; PubMed=2123294;
RA Chavrier P., Vingron M., Sander C., Simons K., Zerial M.;
RA "Molecular cloning of YP1/SEC4-related cDNAs from an epithelial cell

```

```

RT line.";
RL MOI. Cell. Biol. 10:6578-6585(1990).
CC -!- FUNCTION: INVOLVED IN THE TRANSPORT OF PROTEINS BETWEEN
CC THE ENDOSOMES AND THE TRANS GOLGI NETWORK.
CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; X56386; CA39797.1; -.
DR PIR; S36187; S36187.
DR HSP; P01112; IPLJ.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmfrng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small GTP; 1.
KW GTP-binding; Lipoprotein; Prenylation; Protein transport; Golgi stack.
FT NP_BIND 14 21 GTP (BY SIMILARITY).
FT NP_BIND 62 66 GTP (BY SIMILARITY).
FT NP_BIND 124 127 GTP (BY SIMILARITY).
FT DOMAIN 36 44 EFECTOR REGION (BY SIMILARITY).
FT DOMAIN 200 200 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 201 201 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 201 AA; 22810 MW; 65B502C204BEDB72 CRC64;

Query Match 99.5%; Score 1057; DB 1; Length 201;
Best Local Similarity 99.5%; Pred. No. 8.6e-87;
Matches 200; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGKSLFKVILLGDGGVGKSSLMNRYVTKNFKPTQLFHTIGVEFLNKDLEVDGHFVTMQI 60
DB 1 MAGKSLFKVILLGDGGVGKSSLMNRYVTKNFKPTQLFHTIGVEFLNKDLEVDGHFVTMQI 60
QY 61 WDTAGQERFRLTPFPYRGSDCCLLTFSDVSDSOFNLNWKKEFIYYADVKEPSPFFV 120
DB 61 WDTAGQERFRLTPFPYRGSDCCLLTFSDVSDSOFNLNWKKEFIYYADVKEPSPFFV 120
QY 121 ILGNKIDISEROVSTEAQAACWCRDNGDYPYFETSAKDATNVAFAFEAVRVLATEDRSD 180
DB 121 ILGNKIDISEROVSTEAQAACWCRDNGDYPYFETSAKDATNVAFAFEAVRVLATEDRSD 180
QY 181 HLIQTDVNLHRRKPKSSCC 201
DB 181 HLIQTDVNLHRRKPKSSCC 201

RESULT 3
RB9A_RAT STANDARD; PRT; 201 AA.
AC Q99P75;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ras-related protein Rab-9A (Rab-9).
GN RAB9A OR RAB9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Bone;
RX MEDLINE=20047202; PubMed=12051767;
RA Zhao H., Ettala O., Vaananen H.K.;
RA "Intracellular membrane trafficking pathways in bone-resorbing

```



osteoclasts revealed by cloning and subcellular localization studies  
 of small GTP-binding rab proteins.";  
 Biochem. Biophys. Res. Commun. 293:1060-1065(2002).  
 -1- FUNCTION: INVOLVED IN THE TRANSPORT OF PROTEINS BETWEEN  
 THE ENDOSOMES AND THE TRANS GOLGI NETWORK (BY SIMILARITY).  
 -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----

EMBL; AF325692; AAC49586.1; -;  
 HSP; P10114; IKA0.  
 InterPro; IPR003579; GTPase\_Rab.  
 InterPro; IPR001806; Ras\_trnsmfng.  
 InterPro; IPR005225; Small\_GTP.  
 Pfam; PF00071; ras; 1.  
 PRINTS; PR00449; RASTRNSFRMNG.  
 SMART; SM00175; RAB; 1.  
 TIGRfams; TIGR00231; small GTP; 1.  
 GTP-binding; Lipoprotein; Prenylation; Protein transport; Golgi stack.  
 NP\_BIND 14 21  
 GTP (BY SIMILARITY).  
 FT NP\_BIND 62 66  
 GTP (BY SIMILARITY).  
 FT NP\_BIND 124 127  
 GTP (BY SIMILARITY).  
 FT DOMAIN 36 44  
 EFFECTOR REGION (BY SIMILARITY).  
 FT LIPID 200 200  
 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 201 201  
 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 201 AA; 22865 MW; 6E8E3A5AADD656E1 CRC64;

Query Match 97.3%; Score 1033; DB 1; Length 201;  
 Best Local Similarity 95.5%; Pred. No. 1.2e-94;  
 Matches 192; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 MACKSSLFKVLGDDGGVGSLSLMNRYVTNKPTQLPHITIGVEFLNKDLEVDGHFVTMQI 60  
 DB 1 MACKSSLFKVLGDDGGVGSLSLMNRYVTNKPTQLPHITIGVEFLNKDLEVDGHFVTMQI 60  
 QY 61 WDTAGQERFSLRTPFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYADVKEPSPFPV 120  
 DB 61 WDTAGQERFSLRTPFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYADVKEPSPFPV 120  
 QY 121 ILGNKIDISERQVSTEAQAQWCRDNGDYPYFETSAKDNTVAAPAEAVRVLATEDRSD 180  
 DB 121 ILGNKIDISERQVSTEAQAQWCRDNGDYPYFETSAKDNTVAAPAEAVRVLATEDRSD 180  
 QY 181 HLIQTDTVNLHRRKPKPSSSCC 201  
 DB 181 HLIQTDTVNLHRRKPKPSSSCC 201

RESULT 4  
 RB9B\_HUMAN  
 ID RB9B\_HUMAN STANDARD; PRT; 201 AA.  
 AC Q9NP50;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ras-related protein Rab-9B (Rab-9L) (RAB9-like protein).  
 GN RAB9B OR RAB9L.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=20496223; PubMed=11043518;  
 RA Seki N., Azuma T., Yoshikawa T., Masuho Y., Muramatsu M., Saito T.;  
 RT "cDNA cloning of a new member of the Ras superfamily, RAB9-like, on  
 the human chromosome Xq22.1-q22.3 region.";

J. Hum. Genet. 45:318-322(2000).  
 [2] SEQUENCE FROM N.A.  
 RA Bird C.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INVOLVED IN THE TRANSPORT OF PROTEINS BETWEEN  
 THE ENDOSOMES AND THE TRANS GOLGI NETWORK (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: Ubiquitous.  
 CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----

EMBL; AB036693; BAA89542.1; -;  
 EMBL; AL139228; CAB76967.1; -;  
 HSP; P10114; IKA0.  
 Genew; HGNC:14090; RAB9B.  
 DR MIM; 300285; -;  
 DR GO; GO:0005525; P-GTP binding activity; NAS.  
 DR GO; GO:0006895; P-Golgi to endosome transport; NAS.  
 DR InterPro; IPR003579; GTPase\_Rab.  
 DR InterPro; IPR001806; Ras\_trnsmfng.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 PRINTS; PR00449; RASTRNSFRMNG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRfams; TIGR00231; small GTP; 1.  
 GTP-binding; Lipoprotein; Prenylation; Protein transport; Golgi stack.  
 NP\_BIND 14 21  
 GTP (BY SIMILARITY).  
 FT NP\_BIND 62 66  
 GTP (BY SIMILARITY).  
 FT NP\_BIND 124 127  
 GTP (BY SIMILARITY).  
 FT DOMAIN 36 44  
 EFFECTOR REGION (BY SIMILARITY).  
 FT LIPID 200 200  
 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 201 201  
 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 201 AA; 22719 MW; B35EDA8B8358C49C CRC64;

Query Match 79.0%; Score 839; DB 1; Length 201;  
 Best Local Similarity 76.1%; Pred. No. 1.9e-67;  
 Matches 153; Conservative 22; Mismatches 26; Indels 0; Gaps 0;

QY 1 MACKSSLFKVLGDDGGVGSLSLMNRYVTNKPTQLPHITIGVEFLNKDLEVDGHFVTMQI 60  
 DB 1 MACKSSLFKVLGDDGGVGSLSLMNRYVTNKPTQLPHITIGVEFLNKDLEVDGHFVTMQI 60  
 QY 61 WDTAGQERFSLRTPFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYADVKEPSPFPV 120  
 DB 61 WDTAGQERFSLRTPFYRGSDCCLLTFSDVDSQSFQNLNWKKEFIYADVKEPSPFPV 120  
 QY 121 ILGNKIDISERQVSTEAQAQWCRDNGDYPYFETSAKDNTVAAPAEAVRVLATEDRSD 180  
 DB 121 ILGNKIDISERQVSTEAQAQWCRDNGDYPYFETSAKDNTVAAPAEAVRVLATEDRSD 180  
 QY 181 HLIQTDTVNLHRRKPKPSSSCC 201  
 DB 181 HLIQTDTVNLHRRKPKPSSSCC 201

RESULT 5  
 RAB7\_DICDI  
 ID RAB7\_DICDI STANDARD; PRT; 203 AA.  
 AC P36411;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ras-related protein Rab7.  
 GN RAB7.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.

```

OX NCB1_TaxID=44689;
RN SEQUENCE FROM N.A.
RP STRAIN=AX3;
RA Bush J.M. IV, Nolte K., Rodriguez-Paris J., Temesvari L.,
RA Ruscetti T., Steck T., Cardelli J.A.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U02928; AAA80152.1; -.
DR HSP; P01112; IPLJ.
DR DictyDb; DD00070; -.
DR InterPro; IPR003579; GTPase Rab.
DR InterPro; IPR001806; Ras trnsfrmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small_gtp; 1.
KW GTP-binding; Lipoprotein; Prenylation.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 125 128 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 202 202 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 203 203 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 203 AA; 22695 MW; 8984547A7725B521 CRC64;

Query Match
Best Local Similarity 50.4%; Score 535.5; DB 1; Length 203;
Matches 103; Conservative 38; Mismatches 56; Indels 3; Gaps 3;

OY 4 KSSLFVKVILGDGVGKSSLMNRYVYVTKFTQLFHTIGVEFLNKDLEVDGHVFTMQIWD 63
Db 5 KVLVKVILGDGVGKTSLMNQYVKNKFSNQYKATIGADFTKELWDDRVVFTMQIWD 64
OY 64 AGQERFSLRTPYRGSDCLLTFSVDDSQSFQNLNWKKEFYIYADVKEPSPFVILG 123
Db 65 AGQERFQSLGSAFYRGADCCVLVFDVYNNKSFDDLONWRDEFIIQAGPPDPNFPFWLG 124
OY 124 NKIDI--SERQVSTEEAQAWCRDNGDYPYFETSADATNVAAPFAEAVRRVLATEDRSD 182
Db 125 NKIDLENQVRVQKRAASQCKSNIPYFETSAKEINVEQAFQIARNAIKLEDGLVFP 184
OY 183 IQDTVNLHRKPKPS--SSCC 201
Db 185 IPTN-IQVPEPQPAKSGCC 203

RESULT 6
YPTS_CHLRE
ID YPTS_CHLRE STANDARD; PRT; 206 AA.
AC Q39573;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE GTP-binding protein YPTCS.
GN YPTCS
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadales; Chlamydomonas.
OX NCB1_TaxID=3055;
RN SEQUENCE FROM N.A.
RP STRAIN=cw15;

```

```

RX MEDLINE=95309723; PubMed=7789809;
RA Dietmaier W., Fabry S., Huber H., Schmitt R.;
RT "Analysis of a family of ypt genes and their products from
RT Chlamydomonas reinhardtii.";
RL Gene 158:41-50 (1995).
CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
CC TRAFFIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U13170; AAA82728.1; -.
DR PIR; JC4107; JC4107.
DR HSP; P10114; IKAQ.
DR InterPro; IPR003579; GTPase Rab.
DR InterPro; IPR001806; Ras trnsfrmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small_gtp; 1.
KW GTP-binding; Lipoprotein; Prenylation; Protein transport;
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 125 128 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).
FT LIPID 205 205 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 206 206 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 206 AA; 23142 MW; 2F02E469BBDAA0A2 CRC64;

Query Match
Best Local Similarity 50.2%; Score 533; DB 1; Length 206;
Matches 105; Conservative 33; Mismatches 59; Indels 6; Gaps 3;

OY 4 KSSLFVKVILGDGVGKSSLMNRYVYVTKFTQLFHTIGVEFLNKDLEVDGHVFTMQIWD 63
Db 5 KRLKLVILGDGVGKTSLMNQYVKNKFTKEYKATIGADFTLKEIYVDDKKVYMQIWD 64
OY 64 AGQERFSLRTPYRGSDCLLTFSVDDSQSFQNLNWKKEFYIYADVKEPSPFVILG 123
Db 65 AGQERFQSLGSAFYRGADCCVLVFDVYNNKSFDDLONWRDEFIIQAGPPDPNFPFWLG 124
OY 124 NKIDI--SERQVSTEEAQAWCRDNGDYPYFETSADATNVAAPFAEAVRRVLATEDRSD 180
Db 125 NKIDENGSSRQVSEKAKAWCASKGSIYPFETSADKEDINVEAFTCITRNALRNE--KEE 183
OY 181 HLIQDTVNLH--RKPKPSSCC 201
Db 184 ELFPDAVDNNTATQKRGCC 206

RESULT 7
RAB7_PENCL
ID RAB7_PENCL STANDARD; PRT; 206 AA.
AC Q40757;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ras-related protein Rab7 (Possible asposory-associated protein).
OS Pennisetum glabre (Buffelgrass).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACAD clade; Panicoideae; Paniceae; Pennisetum.
OX NCB1_TaxID=35520;
RN SEQUENCE FROM N.A.
RP

```

RC STRAIN-Higgins; TISSUE=Flower;  
 RA Gustine D.L., Hulce D.A., Moyer B.G.;  
 RT "A novel cDNA encoding a rab7-related small GTP-binding protein in  
 RL Penicetum ciliare (buffelgrass).";  
 CC (In) Plant Gene Register PGR95-132.  
 CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U40219; AAA85273.1; --  
 DR HSSP; P01112; 1PLJ  
 DR InterPro; IPR003579; GTPase Rab.  
 DR InterPro; IPR001806; Ras trnsfmg.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTRNSFRMNG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRfams; TIGR00231; small GTP; 1.  
 KW GTP-binding; Lipoprotein; Frenylation; Protein transport.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 125 128 GTP (BY SIMILARITY).  
 FT LPID 205 205 GERANYL-GERANYL (BY SIMILARITY).  
 FT LPID 206 206 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 206 AA; 22996 MW; 4B3B82CC9C65899A CRC64;  
 Query Match 50.0%; Score 531; DB 1; Length 206;  
 Best Local Similarity 49.3%; Pred. No. 4e-40;  
 Matches 100; Conservative 40; Mismatches 57; Indels 6; Gaps 3;  
 QY 4 KSSLFKVILLGDGGVKGSSLMNRYVYTKFTQLFHTIGVEFLNKDLEVDGHFTVMTQWDT 63  
 DB 5 RTLLKVIILGDSGVGKTSLMNQVYVKKFNQYKATIGADFLTKVEQFEDRLFLQLQWDT 64  
 QY 64 AGQERFSLRTPFYRGSDCLLTFSDVDSQSFQNLNWKKEFYIYADVKEPSPFPVILG 123  
 DB 65 AGQERFQSLGVAFYRGADCCVLVDFVTAPNTFKTLDNRDEFLIQAQSPDPENFPVILG 124  
 QY 124 NKIDISERQVSTEEAQCWRDNGDYPYFETSADKATNVAAPFAEAVRVVLATE 180  
 DB 125 NKVDVCGNRRVSEKKAKAWCASKGNIPYFETSAREGNTNVEDAFQCIKVALKNRP-EE 183  
 QY 181 HLIQTDTVNL--HRKPKPSSSCC 201  
 DB 184 ELYVPDIDVVGNGRAQRSSGCC 206  
 RESULT 8  
 RAB7 RAT  
 ID RAB7 RAT STANDARD; PRT; 207 AA.  
 AC P09527;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ras-related protein Rab-7 (RAS-related protein p23) (RAS-related  
 DE protein BRL-RAS).  
 GN RAB7.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Buffalo; TISSUE=Liver;  
 RX MEDLINE=89057527; PubMed=3057452;

RA Bucci C., Franzio R., Chiariotti L., Brown A.L., Rechler M.M.,  
 RA Bruni C.B.;  
 RT "A new member of the ras gene superfamily identified in a rat liver  
 RL cell line.";  
 CC Nucleic Acids Res. 16:9979-9994(1988).  
 CC [2]  
 CC REVISIONS TO N-TERMINUS.  
 RA Bruni C.B.;  
 RA Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Bone;  
 RA Zhao H., Gao L., Vaananen K.H.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: FOUND ON LATE ENDOSOMES (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X12535; CAA31053.1; --  
 DR EMBL; AF286535; AAG00543.1; --  
 DR HSSP; P01112; 1PLJ  
 DR InterPro; IPR003579; GTPase Rab.  
 DR InterPro; IPR001806; Ras trnsfmg.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTRNSFRMNG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRfams; TIGR00231; small GTP; 1.  
 KW GTP-binding; Lipoprotein; Frenylation; Protein transport.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 125 128 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).  
 FT LPID 205 205 GERANYL-GERANYL (BY SIMILARITY).  
 FT LPID 207 207 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 207 AA; 23504 MW; A2AF3B02F672971 CRC64;  
 Query Match 49.9%; Score 529.5; DB 1; Length 207;  
 Best Local Similarity 50.2%; Pred. No. 5.4e-40;  
 Matches 101; Conservative 32; Mismatches 65; Indels 3; Gaps 1;  
 QY 4 KSSLFKVILLGDGGVKGSSLMNRYVYTKFTQLFHTIGVEFLNKDLEVDGHFTVMTQWDT 63  
 DB 5 KVVLLKVIILGDSGVGKTSLMNQVYVKKFNQYKATIGADFLTKVEVVDRLVTMQWDT 64  
 QY 64 AGQERFSLRTPFYRGSDCLLTFSDVDSQSFQNLNWKKEFYIYADVKEPSPFPVILG 123  
 DB 65 AGQERFQSLGVAFYRGADCCVLVDFVTAPNTFKTLDNRDEFLIQAQSPDPENFPVILG 124  
 QY 124 NKIDISERQVSTEEAQCWRDNGDYPYFETSADKATNVAAPFAEAVRVVLATE---DRSD 180  
 DB 125 NKIDLENQVATKKAQAWCYSKNNIPYFETSAREGNTNVEDAFQCIKVALKQETVELYN 184  
 QY 181 HLIQTDTVNLHRKPKPSSSCC 201  
 DB 185 EFPEPKLDKNERAKASAES 205  
 RESULT 9  
 RAB7 CANFA  
 ID RAB7 CANFA STANDARD; PRT; 207 AA.  
 AC P18067;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)  
 Ras-related protein Rab-7.  
 AC RAB7.  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE RAB7.  
 GN RAB7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=90322428; PubMed=2115402;  
 RA Chavrier P., Parton R.G., Hauri H.P., Simons K., Zerial M.;  
 RT "Localization of low molecular weight GTP binding proteins to  
 RT exocytic and endocytic compartments.";  
 RL Cell 62:317-329(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Cocker spaniel;  
 RA Chavrier P., Vingron M., Sander C., Simons K., Zerial M.;  
 RT "Molecular cloning of YPT1/SEC4-related cDNAs from an epithelial cell  
 RT line.";  
 RL Mol. Cell. Biol. 10:6578-6585(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Kim J.Y., Park Y.B.;  
 RT Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: FOUND ON LATE ENDOSOMES.  
 CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; M35522; AAA30890.1; -;  
 CC PIR; B30413; B30413.  
 CC HSP; P01112; 1PLJ.  
 CC InterPro; IPR003579; GTPase Rab.  
 CC InterPro; IPR001806; Ras trnsfrmg.  
 CC InterPro; IPR005225; Small\_GTP.  
 CC Pfam; PF00071; ras; 1.  
 CC PRINTS; PR00449; RASTRNSFRMG.  
 CC SMART; SM00175; RAB; 1.  
 CC TIGRfams; TIGR00231; small GTP; 1.  
 CC NP\_BIND 15 22 GTP (BY SIMILARITY).  
 CC NP\_BIND 63 67 GTP (BY SIMILARITY).  
 CC NP\_BIND 125 128 GTP (BY SIMILARITY).  
 CC DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).  
 CC LIPID 205 205 GERANYL-GERANYL (BY SIMILARITY).  
 CC LIPID 207 207 GERANYL-GERANYL (BY SIMILARITY).  
 CC LIPID 207 207 GERANYL-GERANYL (BY SIMILARITY).  
 CC SEQUENCE 207 AA; 23520 MW; E3AF33B16A67296D CRC64;  
 Query Match 49.8%; Score 529; DB 1; Length 207;  
 Best Local Similarity 56.6%; Pred. No. 6e-40;  
 Matches 98; Conservative 24; Mismatches 51; Indels 0; Gaps 0;  
 QY 4 KSLFLKVLGDSGVKSLMRYVTKPDTQLFHTIGVEFLNKLVDGHHVWQIWD 63  
 Db 5 KKVLKVLGDSGVKSLMRYVTKPDTQLFHTIGVEFLNKLVDGHHVWQIWD 64  
 QY 64 AGQERFSLRTPPYRGSDDLTFSSYDSDQSFQNLNNKKFIYADVKEPSPFPVILG 123  
 Db 65 AGQERFSLRTPPYRGSDDLTFSSYDSDQSFQNLNNKKFIYADVKEPSPFPVILG 124  
 QY 124 NKIDISGVSTEEAQAACWCRDNGDYFETSAKDTNVAARAEAVRVILATE 176  
 Db 125 NKIDLENQVATKRAQAWCYKNNIPYFETSAKEAINVEQAFOTIARNALKE 177  
 RESULT 10  
 RAB7\_HUMAN

ID AC RAB7\_HUMAN STANDARD; PRT; 207 AA.  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ras-related protein Rab-7.  
 GN RAB7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Vitelli R., Chiariello M., Lattero D., Bruni C.B., Bucci C.;  
 RT "Molecular cloning and expression analysis of the human Rab7 GTP-ase  
 RT complementary deoxyribonucleic acid.";  
 RL Biochem. Biophys. Res. Commun. 229:887-890(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=97271569; PubMed=9126495;  
 RA Davies J.P., Cotter P.D., Ioannou Y.A.;  
 RT "Cloning and mapping of human Rab7 and Rab9 cDNA sequences and  
 RT identification of a Rab9 pseudogene.";  
 RL Genomics 41:131-134(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Kim J.Y., Park Y.B.;  
 RT Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: FOUND ON LATE ENDOSOMES (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X93499; CAA63763.1; -;  
 CC EMBL; U44104; AAA86640.1; -;  
 CC EMBL; AF050175; AAD02565.1; -;  
 CC PIR; JCS268; JCS268.  
 CC HSP; P01112; 1PLJ.  
 CC Genew; HGNC:9788; RAB7.  
 CC MIM; 602238; -;  
 CC GO; GO:0005770; C:late endosome; TAS.  
 CC GO; GO:0003928; F:RAB small monomeric GTPase activity; TAS.  
 CC GO; GO:0006897; P:endocytosis; TAS.  
 CC InterPro; IPR003579; GTPase Rab.  
 CC InterPro; IPR001806; Ras trnsfrmg.  
 CC InterPro; IPR005225; Small\_GTP.  
 CC Pfam; PF00071; ras; 1.  
 CC PRINTS; PR00449; RASTRNSFRMG.  
 CC SMART; SM00175; RAB; 1.  
 CC TIGRfams; TIGR00231; small GTP; 1.  
 CC GTP-binding; Lipoprotein; Prenylation; Protein transport.  
 CC NP\_BIND 15 22 GTP (BY SIMILARITY).  
 CC NP\_BIND 63 67 GTP (BY SIMILARITY).  
 CC NP\_BIND 125 128 GTP (BY SIMILARITY).  
 CC DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).  
 CC LIPID 205 205 GERANYL-GERANYL (BY SIMILARITY).  
 CC LIPID 207 207 GERANYL-GERANYL (BY SIMILARITY).  
 CC LIPID 207 207 GERANYL-GERANYL (BY SIMILARITY).  
 CC CONFLICT 47 47 T -> I (IN REF. 3).  
 CC CONFLICT 108 108 I -> V (IN REF. 2).  
 CC CONFLICT 127 127 I -> V (IN REF. 2).  
 CC CONFLICT 180 180 V -> E (IN REF. 3).  
 CC SEQUENCE 207 AA; 23490 MW; A2AF33B16A672971 CRC64;

Query Match 49.8%; Score 529; DB 1; Length 207;  
 Best Local Similarity 56.6%; Pred. No. 6e-40;  
 Matches 98; Conservative 24; Mismatches 51; Indels 0; Gaps 0;

QY 4 KSSLFKVILLGDSGVGKSLMRYVYTNKFDLTQFHTTIGVEFLNKDLEVDGHFVTMQIWD 63  
 DB 5 KKVLLKVIILGDSGVGKSLMRYVYTNKFDLTQFHTTIGVEFLNKDLEVDGHFVTMQIWD 64

QY 64 AGQERFSRLTPFYRGSDCCLLTFEVDSDSQFNLNWKKEFIYYADVKEPSPFVILG 123  
 DB 65 AGQERFSRLTPFYRGSDCCLLTFEVDSDSQFNLNWKKEFIYYADVKEPSPFVILG 124

QY 124 NKIDISERQVSTEEAQAACWCRDNGDYPYFETSADKATNVAAAFEEAVRRVLATE 176  
 DB 125 NKIDLENQVATKRAQAACWCRDNGDYPYFETSADKATNVAAAFEEAVRRVLATE 177

RESULT 11  
 RAB7\_MOUSE  
 ID RAB7\_MOUSE STANDARD; PRT; 207 AA.  
 AC PS1150;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ras-related protein Rab-7.  
 GN RAB7.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96138545; PubMed=8547311;  
 RA Vicelli R., Chiariello M., Bruni C.B., Bucci C.;  
 RT "Cloning and expression analysis of the murine Rab7 cDNA."  
 RL Biochim. Biophys. Acta 1264:268-270(1995).  
 CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: FOUND ON LATE ENDOSOMES (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X89650; CAA61797.1; -  
 DR PIR; S62733; S62733.  
 DR HSP; P01112; IPLJ.  
 DR MGD; MGI:105068; Rab7.  
 DR GO; GO:0005794; C:Golgi apparatus; IDA.  
 DR GO; GO:0005770; C:late endosome; IDA.  
 DR InterPro; IPR003579; GTPase Rab.  
 DR InterPro; IPR001806; Ras trnsfrmg.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTNSFRMG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGR; TIGR00231; small GTP; 1.  
 DR TIGR; TIGR00231; small GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
 FT NP BIND 15 22 GTP (BY SIMILARITY).  
 FT NP BIND 63 67 GTP (BY SIMILARITY).  
 FT NP BIND 125 128 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).  
 FT LIPID 205 205 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 207 207 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 207 AA; A2AB82016A672870 CRC64;

Query Match 49.6%; Score 527; DB 1; Length 207;

Best Local Similarity 56.6%; Pred. No. 9e-40;  
 Matches 98; Conservative 24; Mismatches 51; Indels 0; Gaps 0;

QY 4 KSSLFKVILLGDSGVGKSLMRYVYTNKFDLTQFHTTIGVEFLNKDLEVDGHFVTMQIWD 63  
 DB 5 KKVLLKVIILGDSGVGKSLMRYVYTNKFDLTQFHTTIGVEFLNKDLEVDGHFVTMQIWD 64

QY 64 AGQERFSRLTPFYRGSDCCLLTFEVDSDSQFNLNWKKEFIYYADVKEPSPFVILG 123  
 DB 65 AGQERFSRLTPFYRGSDCCLLTFEVDSDSQFNLNWKKEFIYYADVKEPSPFVILG 124

QY 124 NKIDISERQVSTEEAQAACWCRDNGDYPYFETSADKATNVAAAFEEAVRRVLATE 176  
 DB 125 NKIDLENQVATKRAQAACWCRDNGDYPYFETSADKATNVAAAFEEAVRRVLATE 177

RESULT 12  
 RAB7\_GOSHI  
 ID RAB7\_GOSHI STANDARD; PRT; 207 AA.  
 AC Q9XER8;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ras-related protein Rab7.  
 GN RAB7.  
 OS Gossypium hirsutum (Upland cotton).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Malvales; Malvaceae; Malvoidae; Gossypium.  
 OX NCBI\_TaxID=3635;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Deltapine 62; TISSUE=Etisolated cotyledon;  
 RA Jenkins W.B., Turley R.B., Steele M.;  
 RT "A cDNA clone encoding a Rab7 protein from upland cotton (Gossypium  
 RL hirsutum L.)."  
 CC (In) Plant Gene Register PGR99-064.  
 CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF116243; AAD22451.1; -  
 DR HSP; P36017; LEKO.  
 DR InterPro; IPR003579; GTPase Rab.  
 DR InterPro; IPR001806; Ras trnsfrmg.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTNSFRMG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGR; TIGR00231; small GTP; 1.  
 DR TIGR; TIGR00231; small GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
 FT NP BIND 15 22 GTP (BY SIMILARITY).  
 FT NP BIND 63 67 GTP (BY SIMILARITY).  
 FT NP BIND 125 128 GTP (BY SIMILARITY).  
 FT LIPID 205 205 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 207 207 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 207 AA; 23171 MW; 2453D5D3B0CC617 CRC64;

Query Match 49.3%; Score 524; DB 1; Length 207;  
 Best Local Similarity 51.0%; Pred. No. 1.7e-39;  
 Matches 101; Conservative 34; Mismatches 59; Indels 4; Gaps 2;

QY 7 LFKVILLGDSGVGKSLMRYVYTNKFDLTQFHTTIGVEFLNKDLEVDGHFVTMQIWD 66  
 DB 8 LKVKVILLGDSGVGKSLMRYVYTNKFDLTQFHTTIGVEFLNKDLEVDGHFVTMQIWD 67

QY 67 ERFSLRTPYRGSDCCLLTFSVDDSSQFNLNWKKEFYIYADVKEPSPFPFVILGNKI 126  
 Db 68 ERFSLRTPYRGSDCCLLTFSVDDSSQFNLNWKKEFYIYADVKEPSPFPFVILGNKI 127  
 QY 127 DI---SERQVSTEAQWCRDNGDYPYFETSAKDNTVAAPFEAVRRVLATEDRSD-HL 182  
 Db 128 DVDGNSRVVSEKAKAWCASKGNIPYFETSAKEGPNVDAAFECIARNALKNPEPEIYL 187  
 QY 183 IQDVTNVLHKKPKPSSSC 200  
 Db 188 PETIDVAGGRPQESTGC 205

RESULT 13  
 YPT5 VOLCA STANDARD; PRT; 205 AA.  
 AC P36864;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE GTP-binding protein ypt5.  
 GN YPT5.  
 OS Volvox carteri.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Volvocaceae; Volvox.  
 OX NCBI\_TaxID=3067;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=f. Nagariensis / HK10;  
 RX MEDLINE=94037148; PubMed=8221932;  
 RA Fabry S., Jacobsen A., Huber H., Palme K., Schmitt R.;  
 RT "Structure, expression, and phylogenetic relationships of a family of  
 RT Ypt genes encoding small G-proteins in the green alga Volvox  
 RT carteri";  
 RL Curr. Genet. 24:229-240(1993).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; L08131; AAA34254.1; -;  
 DR PIR; S36368; S36368.  
 DR HSSP; P01112; 1JAH.  
 DR InterPro; IPR003579; GTPase Rab.  
 DR InterPro; IPR001806; Ras trnsfrmg.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTNSFRMG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport;  
 KW Multigene family.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 125 128 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).  
 FT LIPID 204 204 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 205 205 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 205 AA; A0DD8B8764D7E163 CRC64;

Query Match 49.3%; Score 523.5; DB 1; Length 205;  
 Best Local Similarity 50.5%; Pred. No. 1.8e-39;  
 Matches 102; Conservative 34; Mismatches 61; Indels 5; Gaps 3;  
 QY 4 KSLFKVILLGDGGVKGKSLNRYVTKPDFTQLFHTIGVFLNKLQLEVDGHVFTMQIWD 63

Db 5 KRVVLKIILLGDSGVGKTSLNQVQKFKYKATIGADFLTKIEVDKKTWQIWD 64  
 QY 64 AGQERFSLRTPYRGSDCCLLTFSVDDSSQFNLNWKKEFYIYADVKEPSPFPFVILG 123  
 Db 65 AGQERFSLRTPYRGSDCCLLTFSVDDSSQFNLNWKKEFYIYADVKEPSPFPFVILG 124  
 QY 124 NKIDI--SERQVSTEAQWCRDNGDYPYFETSAKDNTVAAPFEAVRRVLATEDRSDH 181  
 Db 125 NKIDVGNVRQVTEKKAKAWCASKGSIYPYFETSAKEDINVERAFTCITRNALRNE-KEEE 183  
 QY 182 LIQDVTNVLHKK--PKPSSSC 201  
 Db 184 LFVPEAVDMNTSATQQRKGCC 205

RESULT 14  
 RAB7\_PEA STANDARD; PRT; 206 AA.  
 ID RAB7\_PEA  
 AC P31022;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ras-related protein Rab7.  
 OS Pisum sativum (Garden pea).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.  
 OX NCBI\_TaxID=3888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=sv. Purpie-podded; TISSUE=Pod;  
 RX MEDLINE=93257636; PubMed=8490139;  
 RA Drew J.E., Bown D., Gatehouse J.A.;  
 RT "Sequence of a novel plant ras-related cDNA from Pisum sativum.";  
 RL Plant Mol. Biol. 21:1195-1199(1993).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X65650; CAA46600.1; -;  
 DR PIR; S33531; S33531.  
 DR HSSP; P36017; 1BK0.  
 DR InterPro; IPR003579; GTPase Rab.  
 DR InterPro; IPR001806; Ras trnsfrmg.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTNSFRMG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 125 128 GTP (BY SIMILARITY).  
 FT LIPID 204 204 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 206 206 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 206 AA; A2126D26DBEBA49 CRC64;

Query Match 49.2%; Score 523; DB 1; Length 206;  
 Best Local Similarity 48.3%; Pred. No. 2e-39;  
 Matches 97; Conservative 42; Mismatches 58; Indels 4; Gaps 2;  
 QY 4 KSLFKVILLGDGGVKGKSLNRYVTKPDFTQLFHTIGVFLNKLQLEVDGHVFTMQIWD 63  
 Db 5 RTLLKVIILLGDSGVGKTSLNQVQKFKYKATIGADFLTKIEVDKKTWQIWD 64

Qy	64	AGCEFRSLRTPPYEGSDCCLLITSVDSDSQFHLNWKKEFYADVKEPSPFPVIG	123
Db	65	AGCEPQSLGVAFYKGADCCCLVVDVNSKFSFLNLRWREFTIOANPSDPENFPFVIG	124
Qy	124	NKIDI--SEROVTEEAQACRONGDYPYPETSAKDATVAAAFEEAVRVLATDRSD	180
Db	125	NKIDIDGNSRVVSEKARACAAKNIPYETSAKEGINVEAFQTIADKSGE-EE	183
Qy	181	HLIQTDVNLHRKPKPSSCC	201
Db	184	ELYLPETIDVGNSSQPRSTGC	204

```

RESULT 15
RAB7_MESCR STANDARD; PRT; 207 AA.
ID RAB7_MESCR STANDARD; PRT; 207 AA.
AC P93267;
DT DT 15-DEC-1998 (Rel. 37, Created)
DT DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE Ras-related protein Rab7A.
OS Mesembryanthemum crystallinum (Common ice plant).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
NCBI_TaxID=3544;
RX [1]
RN
RP SEQUENCE FROM N.A.
RA Michalowski C.B.; Bohnert H.J.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
CC TRAFFIC (BY SIMILARITY).
CC -! SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
CC EMBL; U87142; ABA47557.1; --
CC PIR; T12579; T12579.
CC DR HSSP; P36017; IEXO.
CC DR InterPro; IPR003579; GTPase_Rab.
CC DR InterPro; IPR001806; Ras_trnsfrmg.
CC DR InterPro; IPR005225; Small_GTP.
CC DR Pfam; PF00071; ras; 1.
CC DR PRINTS; PR00449; RASTRNSPRMG.
CC DR SMART; SM00175; RAB; 1.
CC DR TIGRFAMs; TIGR00231; small_GTP; 1.
CC KW GTP-binding; lipidprotein; Prenylation; Protein transport.
CC FT NP_BIND 15 22 GTP (BY SIMILARITY).
CC FT NP_BIND 63 67 GTP (BY SIMILARITY).
CC FT NP_BIND 125 128 GTP (BY SIMILARITY).
CC FT LIPID 205 205 GERANYL-GERANYL (BY SIMILARITY).
CC FT LIPID 207 207 GERANYL-GERANYL (BY SIMILARITY).
CC SQ SEQUENCE 207 AA; 23188 MW; 3B900B895A40D0C7 CRG64;
Query Match 49.0%; Score 520.5; DB 1; Length 207;
Best Local Similarity 48.5%; Pred. No. 3.4e-39;
Matches 98; Conservative 41; Mismatches 58; Indels 5; Gaps 3;
QY 4 KSSLKFKVILLGGGVGKSLMNRVYNKKEDTDLFTHTTGIVVEFLNKDLVDGHFWTQIWD 63
dbb 5 RKLLKIILGDSDVGKTSLSNQFVNKKFSNQYKATIGADFLTKELOFDRLFTLIQWDT 64
QY 64 AQCFERSLRITFFRGSDCLLTFFSVDDDSQSFCNLNWKKEFYIYADVKNKPSFPFVLIG 123
dbb 65 AQCFERFQSLGVAFYRGDCCVLTIDYNNVKSFESLNWRREELIQASPADPNDFPFVLIG 124

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 04:44:31 ; Search time 85 Seconds  
(without alignments)  
610.219 Million cell updates/sec

Title: US-09-988-974-5  
Perfect score: 1062  
Sequence: 1 MAGKSLFKVILLGGVGK.....LIQTDVTLNRKPKPSSCC 201

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1034	97.4	201	11 Q9ROM6	Q9ROM6 mus musculus
2	839	79.0	201	11 Q8BHH2	Q8BHH2 mus musculus
3	531.5	50.0	205	3 Q8TGD9	Q8TGD9 emericella
4	518.5	48.8	209	5 Q23146	Q23146 caenorhabdi
5	518	48.8	206	5 Q9NFG0	Q9NFG0 plasmodium
6	515	48.5	205	10 Q40527	Q40527 nicotiana t
7	514	48.4	206	10 Q40526	Q40526 nicotiana t
8	513.5	48.4	217	10 Q8WOB7	Q8WOB7 oryza sativ
9	513	48.3	206	10 Q40213	Q40213 lotus japon
10	513	48.3	206	10 Q8LS94	Q8LS94 arabidopsis
11	512	48.2	207	5 Q76742	Q76742 drosophila
12	510.5	48.1	256	5 Q9VW6	Q9VW6 drosophila
13	509.5	48.0	205	3 Q9C2L8	Q9C2L8 arabidopsis
14	509	47.9	206	10 Q9C820	Q9C820 arabidopsis
15	508.5	47.9	206	5 Q9NHW8	Q9NHW8 entamoeba h
16	508	47.8	206	10 Q91W76	Q91W76 arabidopsis

17	508	47.8	209	5 Q9U5G8	Q9U5G8 tetrahymena
18	504	47.5	207	10 Q40214	Q40214 lotus japon
19	502.5	47.3	206	5 Q9BLE8	Q9BLE8 entamoeba h
20	501.5	47.2	206	10 Q9X198	Q9X198 arabidopsis
21	500	47.1	206	10 Q948K8	Q948K8 arabidopsis
22	499	47.0	206	10 Q8LGH5	Q8LGH5 arabidopsis
23	498	46.9	204	10 Q40528	Q40528 nicotiana t
24	494.5	46.6	205	10 Q40211	Q40211 lotus japon
25	493	46.4	208	10 Q9S288	Q9S288 arabidopsis
26	481	45.3	212	10 Q9SJ11	Q9SJ11 arabidopsis
27	478	45.0	169	5 O15608	O15608 entamoeba h
28	464.5	43.7	205	10 Q40212	Q40212 lotus japon
29	437	41.1	208	3 Q9HDY0	Q9HDY0 schizosacch
30	433.5	40.8	221	5 Q9XZK0	Q9XZK0 trypanosoma
31	428	40.3	223	5 Q8TEH9	Q8TEH9 leishmania
32	427	40.2	198	10 Q8LQJ2	Q8LQJ2 oryza sativ
33	424	39.9	223	5 Q9N2P5	Q9N2P5 leishmania
34	413	38.9	179	5 Q813B6	Q813B6 plasmodium
35	405	38.1	204	10 Q948K6	Q948K6 arabidopsis
36	398.5	37.5	183	10 Q9FK99	Q9FK99 arabidopsis
37	386.5	36.4	199	11 Q8VEA8	Q8VEA8 mus musculu
38	378.5	35.6	199	4 Q86AH8	Q86AH8 homo sapien
39	356.5	33.6	203	10 Q24112	Q24112 nicotiana p
40	353	33.2	202	10 Q9SEH3	Q9SEH3 arabidopsis
41	353	33.2	221	10 Q23594	Q23594 arabidopsis
42	352.5	33.2	203	10 Q40569	Q40569 nicotiana t
43	352.5	33.2	214	3 Q8WZR7	Q8WZR7 neurospora
44	352	33.1	203	10 Q8RU63	Q8RU63 oryza sativ
45	351	33.1	202	10 Q9FPJ4	Q9FPJ4 arabidopsis

ALIGNMENTS

RESULT 1

Q9ROM6	Q9ROM6	PRELIMINARY;	PRT;	201 AA.
AC	Q9ROM6			
DT	01-MAY-2000 (TREMBlrel. 13, Created)			
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)			
DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)			
DE	Small GTP binding protein (2410064E05R1K protein) (S1 similar to RAB9, member RAS oncogene family) (RAS-related protein RAB-9 homolog).			
GN	RAB9 OR STD 99 GR 2410064E05R1K.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Seki N., Hattori A., Hayashi A., Kozuma S., Muramatsu M., Saito T.;			
RT	"Mouse small GTP binding protein."			
RL	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Embryonic stem cells, and Head;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staali F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Blake J., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,			



RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatauki S.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
[3]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;  
PX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RL 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
DR EMBL; AB027290; BAB84709.1; -;  
DR EMBL; AK017301; BAB30681.1; -;  
DR EMBL; AK010710; BAB27135.1; -;  
DR EMBL; BC008160; AAH08160.1; -;  
DR EMBL; AK032133; BAC27720.1; -;  
DR HSP; P10114; IKAQ.  
DR MGD; MGI:1890695; Rab9.  
DR InterPro; IPR003579; GTPase\_Rab.  
DR InterPro; IPR001806; Ras\_trnsfrmng.  
DR InterPro; IPR005225; Small\_GTP.  
DR Pfam; PF00071; ras; 1.  
DR PRINTS; PR00449; RASTRNSFRMNG.  
DR SMART; SM00175; RAB; 1.  
DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
DR GTP-binding; Lipoprotein.  
KW GTP-binding; Lipoprotein.  
SQ SEQUENCE 201 AA; 22909 MW; 43F4908AA67D2C9A CRC64;  
  
Query Match 97.4%; Score 1034; DB 11; Length 201;  
Best Local Similarity 95.5%; Pred. No. 6.7e-90;  
Matches 192; Conservative 7; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 MAGKSLFKVILLGDGGVKGKSLMNRVYTNKFDQTFHTIGVEFLNKLQLEVDGHFVTWQI 60  
DB 1 MAGKSLFKVILLGDGGVKGKSLMNRVYTNKFDQTFHTIGVEFLNKLQLEVDGHFVTWQI 60  
  
QY 61 WDTAGQERFSLRTPFYRGSDCLLTFSVDDSQSFQNLNWKKEFYIYADVKEPSPFFV 120  
DB 61 WDTAGQERFSLRTPFYRGSDCLLTFSVDDSQSFQNLNWKKEFYIYADVKEPSPFFV 120  
  
QY 121 ILGNKIDISERQVSTEEAQAWCRDNGDYPYFETSAKDATNVAAPFAEVRVLATEDRSD 180  
DB 121 ILGNKIDISERQVSTEEAQAWCRDNGDYPYFETSAKDATNVAAPFAEVRVLATEDRSD 180  
  
QY 181 HLIQTDVNLHRRKPKSPSSCC 201  
DB 181 HLIQTDVNLHRRKPKSPSSCC 201  
  
RESULT 2  
Q8BHH2 PRELIMINARY; PRT; 201 AA.  
ID Q8BHH2  
AC Q8BHH2;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE RAS-related protein RAB-9L.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=diencephalon, and Spinal cord;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RL 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
DR EMBL; AK034442; BAC28710.1; -;  
DR EMBL; AK049693; BAC33876.1; -;  
SQ SEQUENCE 201 AA; 22704 MW; 1E28B18F8F8DDDFD CRC64;  
  
Query Match 79.0%; Score 839; DB 11; Length 201;  
Best Local Similarity 76.1%; Pred. No. 1.9e-71;  
Matches 153; Conservative 24; Mismatches 24; Indels 0; Gaps 0;  
  
QY 1 MAGKSLFKVILLGDGGVKGKSLMNRVYTNKFDQTFHTIGVEFLNKLQLEVDGHFVTWQI 60  
DB 1 MSGKSLLLKVVILLGDGGVKGKSLMNRVYTNKFDQTFHTIGVEFLNKLQLEVDGHFVTWQI 60  
  
QY 61 WDTAGQERFSLRTPFYRGSDCLLTFSVDDSQSFQNLNWKKEFYIYADVKEPSPFFV 120  
DB 61 WDTAGQERFSLRTPFYRGSDCLLTFSVDDSQSFQNLNWKKEFYIYADVKEPSPFFV 120  
  
QY 121 ILGNKIDISERQVSTEEAQAWCRDNGDYPYFETSAKDATNVAAPFAEVRVLATEDRSD 180  
DB 121 VLGKVKDEKQVTEEAQAWCRDNGDYPYFETSAKDATNVAAPFAEVRVLATEDRSD 180  
  
QY 181 HLIQTDVNLHRRKPKSPSSCC 201  
DB 181 HMLGHTIDLNSGSKASSCC 201  
  
RESULT 3  
Q8TGD9 PRELIMINARY; PRT; 205 AA.  
ID Q8TGD9  
AC Q8TGD9;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Small GTPase AvasA.  
GN AVAA.  
OS Emericella nidulans (Aspergillus nidulans).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; Emericella.  
OX NCBI\_TaxID=162425;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A26;  
RA Ohsumi K., Nakajima H., Kitamoto K.;  
RT "Cloning and characterization of Aspergillus nidulans avaa gene  
RT encoding a homologue of Rab7/Ypt7 GTPase involved in vacuolar  
RT biogenesis";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
DR EMBL; AB072431; BAB86862.1; -;  
DR HSP; P05713; 3RAB.  
DR InterPro; IPR003579; GTPase\_Rab.  
DR InterPro; IPR001806; Ras\_trnsfrmng.  
DR InterPro; IPR002078; Sig54\_interact.  
DR InterPro; IPR005225; Small\_GTP.  
DR Pfam; PF00071; ras; 1.  
DR SMART; SM00175; RAB; 1.  
DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
DR PROSITE; PS00675; SIGMAS4\_INTERACT\_1; 1.  
DR GTP-binding; Lipoprotein.  
KW GTP-binding; Lipoprotein.  
SQ SEQUENCE 205 AA; 22941 MW; 0E927E2B76F6BB32 CRC64;  
  
Query Match 50.0%; Score 531.5; DB 3; Length 205;  
Best Local Similarity 50.2%; Pred. No. 2.5e-42;  
Matches 101; Conservative 37; Mismatches 60; Indels 3; Gaps 2;  
  
QY 4 KSSLFKVILLGDGGVKGKSLMNRVYTNKFDQTFHTIGVEFLNKLQLEVDGHFVTWQI 63  
DB 5 KVMVLKVIILGDGGVKGKSLMNRVYTNKFDQTFHTIGVEFLNKLQLEVDGHFVTWQI 64  
  
QY 64 AQGERFSLRTPFYRGSDCLLTFSVDDSQSFQNLNWKKEFYIYADVKEPSPFFVILG 123

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatauki S.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
[3]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;  
PX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RL 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
DR EMBL; AB027290; BAB84709.1; -;  
DR EMBL; AK017301; BAB30681.1; -;  
DR EMBL; AK010710; BAB27135.1; -;  
DR EMBL; BC008160; AAH08160.1; -;  
DR EMBL; AK032133; BAC27720.1; -;  
DR HSP; P10114; IKAQ.  
DR MGD; MGI:1890695; Rab9.  
DR InterPro; IPR003579; GTPase\_Rab.  
DR InterPro; IPR001806; Ras\_trnsfrmng.  
DR InterPro; IPR005225; Small\_GTP.  
DR Pfam; PF00071; ras; 1.  
DR PRINTS; PR00449; RASTRNSFRMNG.  
DR SMART; SM00175; RAB; 1.  
DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
DR GTP-binding; Lipoprotein.  
KW GTP-binding; Lipoprotein.  
SQ SEQUENCE 201 AA; 22909 MW; 43F4908AA67D2C9A CRC64;  
  
Query Match 97.4%; Score 1034; DB 11; Length 201;  
Best Local Similarity 95.5%; Pred. No. 6.7e-90;  
Matches 192; Conservative 7; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 MAGKSLFKVILLGDGGVKGKSLMNRVYTNKFDQTFHTIGVEFLNKLQLEVDGHFVTWQI 60  
DB 1 MAGKSLFKVILLGDGGVKGKSLMNRVYTNKFDQTFHTIGVEFLNKLQLEVDGHFVTWQI 60  
  
QY 61 WDTAGQERFSLRTPFYRGSDCLLTFSVDDSQSFQNLNWKKEFYIYADVKEPSPFFV 120  
DB 61 WDTAGQERFSLRTPFYRGSDCLLTFSVDDSQSFQNLNWKKEFYIYADVKEPSPFFV 120  
  
QY 121 ILGNKIDISERQVSTEEAQAWCRDNGDYPYFETSAKDATNVAAPFAEVRVLATEDRSD 180  
DB 121 ILGNKIDISERQVSTEEAQAWCRDNGDYPYFETSAKDATNVAAPFAEVRVLATEDRSD 180  
  
QY 181 HLIQTDVNLHRRKPKSPSSCC 201  
DB 181 HLIQTDVNLHRRKPKSPSSCC 201  
  
RESULT 2  
Q8BHH2 PRELIMINARY; PRT; 201 AA.  
ID Q8BHH2  
AC Q8BHH2;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE RAS-related protein RAB-9L.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=diencephalon, and Spinal cord;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,

Db 65 AGQERFQSLGVAFYRGADCCVYVYNNKSFALDSWRDEFLIQASPRDPSPFPVVG 124  
Qy 124 NKIDI--SERQVSTEEAQCNRNGDYPPYFETSAKQATNVAFAFEAVRRVLATEDRSDH 181  
Db 125 NKIDMEESKMISSKRAMTFCQSGNTPYFETSAKEAVNVEQAFVIARSALAQEAEAY 184  
Qy 182 -LIQTDTVNLHRKPKPSSSC 201  
Db 185 GGDYTDPIINHDTIRGCGAC 205

## RESULT 4

Q23146 PRELIMINARY; PRT; 209 AA.  
AC Q23146;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE W03C9.3 protein.  
GN W03C9.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gregory J., Ainscough R.;  
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology."  
RL Science 282:2012-2018 (1998).  
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
DR EMBL; Z66516; CAA91357.1; --  
DR HSSP; P01112; 1PLJ.  
DR WormPep; W03C9.3; CE03777.  
DR InterPro; IPR003579; GTPase Rab.  
DR InterPro; IPR001806; Ras trnsfrmg.  
DR InterPro; IPR002078; Sig54 interact.  
DR Pfam; PF00071; ras; 1.  
DR PRINTS; PR00449; RASTNSFRMG.  
DR SMART; SM00175; RAB; 1.  
DR TIGRFAMS; TIGR00231; small GTP; 1.  
DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
KW GTP-binding; Lipoprotein.  
SQ SEQUENCE 209 AA; 23429 MW; 1797B3E7B887F255 CRC64;

Query Match 48.8%; Score 518.5; DB 5; Length 209;  
Best Local Similarity 48.8%; Pred. No. 4.4e-41;  
Matches 101; Conservative 43; Mismatches 56; Indels 7; Gaps 4;

Qy 1 MAG--KSLFKVILLGDGGVGKSSLMNRYVTKFDTLQFLHTIGVFLNKDLVDGHVFM 58  
Db 1 MSGRKALKVILLGDSGVGKTSLMNQVNNKFTNQYKATIGADFLTDVNDRTVTL 60  
Qy 59 QIWDTAGQERFSLRTPFVRGSDCLLTFSVDDSSQFQNLNWKKEFYIYADVKEPESFP 118  
Db 61 QIWDTAGQERFQSLGVAFYRGADCCVLAFLDVTNAAAFKSLDSWRDEFLIQASPRDPDHP 120  
Qy 119 FVILGNKIDI--SRQVSTEEAQCNRNGDYPPYFETSAKQATNVAFAFEAVRRVLATE- 176  
Db 121 FVLGNKVDLESQRAVSQKRAQSCQTKGNIPIYEVSAKEALNVAFLAIARDALAES 180  
Qy 177 -DRSDHLIQTDTNL--HRKPKPSSSC 200  
Db 181 QETNDPFPDQIRLNPQNQNSGC 207

## RESULT 5

Q9NFGO PRELIMINARY; PRT; 206 AA.  
AC Q9NFGO;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative Rab7 GTPase.  
GN RAB7.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=21036610; PubMed=11163452;  
RA Spielmann T., Beck H.P.;  
RT "Analysis of stage-specific transcription in Plasmodium falciparum  
reveals a set of genes exclusively transcribed in ring stage  
parasites."  
RL Mol. Biochem. Parasitol. 111:453-458 (2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RA Quevillon E., Spielmann T., Langsley G.;  
RT "The rab gene family of Plasmodium falciparum";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
DR EMBL; AJ290938; CAB92946.2; --  
DR HSSP; P10114; 1XAO.  
DR InterPro; IPR003579; GTPase Rab.  
DR InterPro; IPR001806; Ras trnsfrmg.  
DR InterPro; IPR002078; Sig54 interact.  
DR InterPro; IPR005225; Small\_GTP.  
DR Pfam; PF00071; ras; 1.  
DR PRINTS; PS00449; RASTNSFRMG.  
DR SMART; SM00175; RAB; 1.  
DR TIGRFAMS; TIGR00231; small GTP; 1.  
DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
KW GTP-binding; Lipoprotein.  
SQ SEQUENCE 206 AA; 23788 MW; E46BD3A7460084D5 CRC64;

Query Match 48.8%; Score 518; DB 5; Length 206;  
Best Local Similarity 47.5%; Pred. No. 4.8e-41;  
Matches 96; Conservative 42; Mismatches 60; Indels 4; Gaps 2;

Qy 4 KSLFKVILLGDGGVGKSSLMNRYVTKFDTLQFLHTIGVFLNKDLVDGHVFMQIWD 63  
Db 5 KRTILKVIILLGDSGVGKTSLMNQVNNKFTNQYKATIGADFLTKETIVDNEQITMQIWD 64  
Qy 64 AQQERFSLRTPFVRGSDCLLTFSVDDSSQFQNLNWKKEFYIYADVKEPESFPVILG 123  
Db 65 AQQERFQSLGVAFYRGADCCVLDLTNYKTSLESWKDEFLIQASPKDPENFPFVIIG 124  
Qy 124 NKID-ISRQVSTEEAQCNRNGDYPPYFETSAKQATNVAFAFEAVRRVLATEDRSDHL 182  
Db 125 NKYDETNRKRVQSLKVLQWCKSNNTIPYFETSAKNAINVDAQDEIARAKMKQEHQEQI 184  
Qy 183 IQDTVNLHRKPKP---SSSCC 201  
Db 185 YLPETPALNNQSEQKMKSRCC 206

## RESULT 6

Q40527 PRELIMINARY; PRT; 205 AA.  
AC Q40527;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Nt-rab7b protein.  
GN NT-RAB7B.  
OS Nicotiana tabacum (Common tobacco).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SRI;  
 RX MEDLINE=95303981; PubMed=7784525;  
 RA Haizel T., Merkle T., Turck F., Nagy F.;  
 RT "Characterization of membrane-bound small GTP-binding proteins from  
 RT Nicotiana tabacum."  
 RL Plant Physiol. 108:59-67(1995).  
 CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 DR EMBL; L29275; AAA74119.1; -.  
 DR HSSP; P36017; 1EK0.  
 DR InterPro; IPR003579; GTPase\_Rab.  
 DR InterPro; IPR001806; Ras\_trnsfrmng.  
 DR InterPro; IPR002078; Sig54\_interact.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTNSFRMNG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRFAMS; TIGR00231; small GTP; 1.  
 DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
 KW GTP-binding; Lipoprotein.  
 SQ SEQUENCE 205 AA; 22744 MW; 596A6F1BA68A7D1B CRC64;

Query Match 48.5%; Score 515; DB 10; Length 205;  
 Best Local Similarity 46.6%; Pred. No. 9.2e-41;  
 Matches 95; Conservative 44; Mismatches 61; Indels 4; Gaps 2;  
 QY 1 MAGKSLFKVILLGGGVGKSLMRYVYTNKFDQTLFHTIGVEFLNKDLEVDGHVFTMQI 60  
 DB 1 MPSPANVLKVIILGDSGVGKTSMLNQVYVNRKFSNQYKATIGADFLTKVEQFEDRLFTLI 60  
 QY 61 WTAGQRRFRSLRTPFYRGSDCCLLTFSDVDSQFNLNWKKEFIYYADVKEPSPFV 120  
 DB 61 WTAGQRRFQSLGVAFYRGADCCVLVYVNRKFSFENLNWREELIQASPSDENPFV 120  
 QY 121 ILGNKIDI---SERQVSTEEAAQWCRDNGDYYPFETSAKDATNVAAAFEEAVRVLAT 177  
 DB 121 VLGNKVIDDGNRSRVSEKKVAKWASKGNIPYFETSAKEGTFVEAFQCIKNAKSGE 180  
 QY 178 RSDHLIQTDTVNLHRKPKPSSCC 201  
 DB 181 -EEIYLPDTLDVGVSSQPRIGCC 203

RESULT 7  
 Q40526 PRELIMINARY; PRT; 206 AA.  
 ID Q40526  
 AC Q40526  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Nt-rab7a protein.  
 GN NT-RAB7A.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SRI;  
 RX MEDLINE=95303981; PubMed=7784525;  
 RA Haizel T., Merkle T., Turck F., Nagy F.;  
 RT "Characterization of membrane-bound small GTP-binding proteins from  
 RT Nicotiana tabacum."  
 RL Plant Physiol. 108:59-67(1995).  
 CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 DR EMBL; L29274; AAA74118.1; -.  
 DR HSSP; P36017; 1EK0.

DR InterPro; IPR003579; GTPase\_Rab.  
 DR InterPro; IPR001806; Ras\_trnsfrmng.  
 DR InterPro; IPR002078; Sig54\_interact.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTNSFRMNG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRFAMS; TIGR00231; small GTP; 1.  
 DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
 KW GTP-binding; Lipoprotein.  
 SQ SEQUENCE 206 AA; 23104 MW; 59CDBCDBDA4794294 CRC64;  
 Query Match 48.4%; Score 514; DB 10; Length 206;  
 Best Local Similarity 47.8%; Pred. No. 1.2e-40;  
 Matches 97; Conservative 40; Mismatches 62; Indels 4; Gaps 2;  
 QY 2 AGKSLFKVILLGGGVGKSLMRYVYTNKFDQTLFHTIGVEFLNKDLEVDGHVFTMQI 61  
 DB 3 ARRRMLKVIILGDSGVGKTSMLNQVYVNRKFSNQYKATIGADFLTKVEQFEDRLFTLI 62  
 QY 62 DTAGQRRFRSLRTPFYRGSDCCLLTFSDVDSQFNLNWKKEFIYYADVKEPSPFV 121  
 DB 63 DTAGQRRFQSLGVAFYRGADCCVLVYVNRKFSFENLNWREELIQASPSDENPFV 122  
 QY 122 LGNKIDI---SERQVSTEEAAQWCRDNGDYYPFETSAKDATNVAAAFEEAVRVLAT 178  
 DB 123 LGNKIDVDGNSRVSEKKVAKWASKGNIPYFETSAKEGTFVEAFQCIKNAKNEP- 181  
 QY 179 SDHLIQTDTVNLHRKPKPSSCC 201  
 DB 182 EDEIYLPDTLDVAGSQSRSTGC 204

RESULT 8  
 Q8W0B7 PRELIMINARY; PRT; 217 AA.  
 ID Q8W0B7  
 AC Q8W0B7  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE RAS-related GTP-binding protein Rab7 family.  
 GN P0452F10.6.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone:P0452F10."  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 DR EMBL; AP003434; BAB78669.1; -.  
 DR HSSP; P05713; 3RAB.  
 DR Gramene; Q8W0B7; -.  
 DR InterPro; IPR003579; GTPase\_Rab.  
 DR InterPro; IPR001806; Ras\_trnsfrmng.  
 DR InterPro; IPR002078; Sig54\_interact.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTNSFRMNG.  
 DR SMART; SM00175; RAB; 1.  
 DR TIGRFAMS; TIGR00231; small GTP; 1.  
 DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
 KW GTP-binding; Lipoprotein.  
 SQ SEQUENCE 217 AA; 24361 MW; 7F32441A2F08183B CRC64;

Query Match 48.4%; Score 513.5; DB 10; Length 217;  
 Best Local Similarity 52.5%; Pred. No. 1.4e-40;  
 Matches 96; Conservative 34; Mismatches 48; Indels 5; Gaps 2;

```

QY 7 LFKVILLGGVGKSSLMNRYVTKNKEPTQLFHTIGVEFLNKLQLEVDGHFVTMOIWDTAGQ 66
DB 8 LLKVILLGDSGVGKISLMNQVNVKFNQYKATIGADFLLTKEVQIDRDLFTLQIWDTAGQ 67
QY 67 ERPSLRTPTFYRGSDCCLLTFSVDDSSQFNLNWKKEFYIYADVKEPESFPFVILGNKI 126
DB 68 ERFQSLGVAFYRGADCCVLVDVNVTKSFERLNSWREELFIQASPSDPENFPFVILGNKI 127
QY 127 DI---SRRQVSTEAQACRNDGYPYFETSAKDNTVAAPFAEAVRVLATEDRSD--H 181
DB 128 DVDGNGERTVSEKKAKWCAASKGNIPYFETSAKEGNEVAEFCEIARNAIKNPESEIIF 187
QY 182 LIQ 184
DB 188 LIQ 190

RESULT 9
Q40213 PRELIMINARY; PRT; 206 AA.
AC Q40213
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RAB7C.
GN RAB7C.
OS Lotus japonicus.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.
OX NCBI_TaxID=34305;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Root nodules;
RX MEDLINE=97231679; PubMed=9076991;
RA Borg S., Brandstrup B., Jensen T.J., Poulsen C.;
RT "Identification of new protein species among 33 different small GTP-
RT binding proteins encoded by cDNAs from Lotus japonicus, and expression
RT of corresponding mRNAs in developing root nodules.";
RL Plant J. 11:237-250(1997).
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL; Z73942; CAA98170.1; -.
DR HSSP; P36017; 1EKO.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trsfmng.
DR InterPro; IPR002078; Sig54_interact.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRfam; TIGR00231; small_GTP; 1.
DR PROSITE; PS00675; SIGMAS4_INTERACT_1; 1.
KW GTP-binding; Lipoprotein.
SQ SEQUENCE 206 AA; 23025 MW; F2453D6B1705177F CRC64;

Query Match
Best Local Similarity 48.3%; Score 513; DB 10; Length 206;
Matches 96; Conservative 40; Mismatches 61; Indels 4; Gaps 2;

QY 4 KSLFKVILLGGVGKSSLMNRYVTKNKEPTQLFHTIGVEFLNKLQLEVDGHFVTMOIWD 63
DB 5 RRTLLKVIILGDSGVGKISLMNQVNVKFNQYKATIGADFLLTKEVQIDRDLFTLQIWD 64
QY 64 AQGERFSLRTPTFYRGSDCCLLTFSVDDSSQFNLNWKKEFYIYADVKEPESFPFVILG 123
DB 65 AQGERFQSLGVAFYRGADCCVLVDVNVTKSFERLNSWREELFIQASPSDPENFPFVILG 124
QY 124 NKIDI---SERQVSTEAQACRNDGYPYFETSAKDNTVAAPFAEAVRVLATEDRSD 180
DB 125 NKIDIDGNSRVVSEKKARAWCASKGNIPYFETSAKGINVEEAFQCIKALKSGE-EE 183
QY 181 HLIQTDVTVNLNHRKPKPSSCC 201

```

```

DB 184 ELYLPDITDVGNSSQORASGC 204

RESULT 10
Q9LS94 PRELIMINARY; PRT; 206 AA.
AC Q9LS94
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RAS-related GTP-binding protein (Atrab71) (Putative RAS-related GTP-
DE binding protein).
GN ATRAB71 OR MVE11.21/AT3G18820.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Ueda T., Wada Y., Nakano A.;
RT "Rab7 homologs in Arabidopsis thaliana.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene MVE11.21/AT3G18820 (GI:9293907)";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yanamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Full length cDNA of gene MVE11.21/AT3G18820 (GI:9293907)";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL; AB026654; BAB01810.1; -.
DR EMBL; AB071846; BAB68371.1; -.
DR EMBL; AY035137; AAK59641.1; -.
DR EMBL; AY059072; AAL15178.1; -.
DR HSSP; P36017; 1EKO.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trsfmng.
DR InterPro; IPR002078; Sig54_interact.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.

```



```

RESULT 12
Q9V1W6 PRELIMINARY; PRT; 256 AA.
ID Q9V1W6
AC Q9V1W6
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE CG9994 protein (REI7845p).
GN CG9994
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKLEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA April J.P., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bekman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brostier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foelel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Drenek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacle J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celinker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL; AE003662; AAF53798.1; -.
DR EMBL; AY071139; AAL48761.1; -.
DR HSSP; P05713; 3RAB.
DR FlyBase; FBgn0032782; CG9994.

DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmfmg.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small_GTP; 1.
KW GTP-binding; Lipoprotein.
SQ SEQUENCE 256 AA; 28467 MW; 100B1A527984DC67 CRC64;

Query Match 48.1%; Score 510.5; DB 5; Length 256;
Best Local Similarity 51.5%; Pred. No. 3.3e-40;
Matches 101; Conservative 35; Mismatches 49; Indels 11; Gaps 4;

QY 4 KSSLFKVILLGDGGVGKSSLMNRYVTKFTDLPHTTGVFLNKLDEVDGHFTVMQIWDT 63
DB 9 KSKLLKVVILGDGGVGKSSALTRFVAVNRYENNFTTIGVFLNKLDEVDGHFTVMQIWDT 68
QY 64 AGQERFRSLRTPFYRGSDCLLTPTSDSOSFQNLNWKKEFIYADVKEPEGFPFVILG 123
DB 69 AGQERFRSLRTPFYRGSDICLLCYALDDRLSKGLGWRNEFLNYADV-DQDKFPFVILG 127
QY 124 NKIDI--SERQVSTEEAQCACRDNDYYPETSAKADINVAAPFEAEVRR-----VLATE 176
DB 128 NKNDIPAKQKQVSDAVQWCAEQKACHTETSSKAATNTVDAFVLGLQRWHEMCVAAE 187
QY 177 DRSDHLIQTDTVNLHR 192
DB 188 ELRQH---GDTIDLTR 200

RESULT 13
Q9C2L8 PRELIMINARY; PRT; 205 AA.
ID Q9C2L8
AC Q9C2L8
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Probable GTPase Rab7 protein.
GN 17B5.300.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariaceae; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schultze U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL; AL513467; CAC28856.1; -.
DR HSSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmfmg.
DR InterPro; IPR002078; Sig54_interact.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small_GTP; 1.
DR PROSITE; PS00675; SIGMAS4_INTERACT_1; 1.
KW GTP-binding; Lipoprotein.
SQ SEQUENCE 205 AA; 23128 MW; EFC8EBB8B89528 CRC64;

Query Match 48.0%; Score 509.5; DB 3; Length 205;
Best Local Similarity 48.8%; Pred. No. 3.1e-40;
Matches 98; Conservative 36; Mismatches 64; Indels 3; Gaps 2;

QY 4 KSSLFKVILLGDGGVGKSSLMNRYVTKFTDLPHTTGVFLNKLDEVDGHFTVMQIWDT 63

```

```

Db 5 KVLKLVILGDSGVGKSLNQVNVKFSYATIGADFLFREVLVDDQVMTQMDT 64
Qy 64 AGQRFSLRTPFYRGSDCLLTFSDVDSQFQNLNWKKEFYIYADVKPESPFPVILG 123
Db 65 AGQRFQSLGVAFYRGADCCVLYDVNNSKFDALDSWRDEFLIQASPRDPDNFPVILG 124
Qy 124 NKIDI--SERQVSTEEAQCWCRNGDYPYFETSKDQATNVAAPFEAVRVLATERSDH 181
Db 125 NKIDMESKRVISYRATWTCQSGKNIPYFETSAKEAINVEQAFVIARNALMQEESBF 184
Qy 182 LIQ-TDVTNLHRKPKPSSCC 201
Db 185 SGDFOBDINHIENDRGAC 205

RESULT 14
Q9C820 PRELIMINARY; PRT; 206 AA.
AC Q9C820;
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE GTP-binding protein RAB7D, putative (Acrab72) (Putative GTP-binding protein RAB7D).
GN F19K6.10 OR ATRAB72 OR ATIG52280.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Ericaceae; Rosaceae;
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A., Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";
RL Nature 408:816-820 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Ueda T., Wada Y., Nakano A.;
RT "Rab7 homologs in Arabidopsis thaliana.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M., Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RA "Full Length cDNA of gene At1G52280 (GI:15218194).";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.

```

```

RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M., Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RA "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC037424; AAG51552.1; -
DR EMBL; AB071847; BAB68372.1; -
DR EMBL; AY074361; AAL67057.1; -
DR EMBL; AY086407; AAM20047.1; -
DR HSP; F36017; IEKO.
DR InterPro; IPR001806; Ras_trnsfrmg.
DR InterPro; IPR002078; Sig54_interact.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR TIGRPFMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW GTP-binding.
SQ SEQUENCE 206 AA; 23069 MW; 6E31BB86CD75CB7F CRC64;
Query Match 47.9%; Score 509; DB 10; Length 206;
Best Local Similarity 48.5%; Pred. No. 3.4e-40;
Matches 96; Conservative 37; Mismatches 61; Indels 4; Gaps 2;
Qy 7 LFKVILLGDGVGKSLNRYVTKPDTQFLHTTIGVEFLNKLVDGHHFVMTQIWDTAGQ 66
Db 8 LLKVIILLGDSGVGKSLNRYVTKPDTQFLHTTIGVEFLNKLVDGHHFVMTQIWDTAGQ 67
Qy 67 ERFRSLRTPFYRGSDCLLTFSDVDSQFQNLNWKKEFYIYADVKPESPFPVILGKI 126
Db 68 ERFSGLVAFYRGADCCVLYDVNNSKFDALDSWRDEFLIQASPRDPDNFPVILGKI 127
Qy 127 DI---SERQVSTEEAQCWCRNGDYPYFETSKDQATNVAAPFEAVRVLATERSDHLI 183
Db 128 DVDGKGRVSEKKAKAWCASKNIPYFETSAKEGNVDAFAECITKNAFKNPEPEEPYL 187
Qy 184 QTDVTNLHRKPKPSSCC 201
Db 188 -PDTIDVAGGQQQRSTGC 204

ID Q9NHWS PRELIMINARY; PRT; 206 AA.
AC Q9NHWS;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Rab7-like GTPase (Fragment).
GN RAB7.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RA Temesvari L.A.;
RT "A Rab7-like GTPase of Entamoeba histolytica.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAS SUBFAMILY.
DR EMBL; AF218311; AAF32317.1; -
DR HSP; F36017; IEKO.
DR InterPro; IPR003579; GTPase Rab.
DR InterPro; IPR001806; Ras_trnsfrmg.
DR InterPro; IPR002078; Sig54_interact.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.

```

Search completed: February 10, 2004, 04:58:12  
Job time : 90 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 23:58:36 ; Search time 4588 Seconds  
(without alignments)  
10477.073 Million cell updates/sec

Title: US-09-988-974-6  
Perfect score: 1175  
Sequence: 1 CACTGTGATGAACACTTTT.....GTAAAAA.....AAAAA 1175

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rnd.\*
- 36: em\_htg\_rnd.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1175	100.0	1175	6	AR034083	AR034083 Sequence
2	1100	93.6	213863	9	AC003037	AC003037 Homo sapi
3	993.2	84.5	1106	9	BC017265	BC017265 Homo sapi
C 4	818.2	69.6	143872	9	AC091987	AC091987 Homo sapi
5	713.6	60.7	1001	4	CFRAB9	X58385 Canine rab9
C 6	696.6	59.3	130407	2	AC084354	AC084354 Homo sapi
7	696.6	59.3	156357	9	AC022486	AC022486 Homo sapi
C 8	686.2	58.4	183293	9	AC142348	AC142348 Pan trogl
C 9	682.8	58.1	161879	9	AC017005	AC017005 Homo sapi
10	682.8	58.1	166436	9	AC016752	AC016752 Homo sapi
C 11	682.8	58.1	182083	9	AC007965	AC007965 Homo sapi
12	664.6	56.6	1251	10	BC008160	BC008160 Mus muscu
13	660.2	56.2	1276	10	AB027290	AB027290 Mus muscu
C 14	656.6	55.9	201642	10	AL672174	AL672174 Mouse DNA
15	646.6	55.0	184474	10	AL672047	AL672047 Mouse DNA
16	646.6	55.0	204031	2	AC093354	AC093354 Mus muscu
17	638.2	54.3	208391	2	AC130009	AC130009 Rattus no
C 18	638.2	54.3	217558	2	AC111489	AC111489 Rattus no
19	632.4	53.8	1286	10	AF325692	AF325692 Rattus no
20	604.4	51.4	606	6	AR270519	AR270519 Sequence
21	604.4	51.4	606	9	AF498944	AF498944 Homo sapi
22	604.4	51.4	606	9	HSU44103	U44103 Human small
23	492	41.9	64450	9	AC073893	AC073893 Homo sapi
C 24	492	41.9	145383	9	AC010682	AC010682 Homo sapi
25	492	41.9	149794	2	AC007273	AC007273 Homo sapi
C 26	439.6	37.4	487	6	AR237295	AR237295 Sequence
C 27	439.6	37.4	487	6	AR260788	AR260788 Sequence
C 28	439.6	37.4	487	6	AR278319	AR278319 Sequence
C 29	439.6	37.4	487	6	AX106425	AX106425 Sequence
C 30	439.6	37.4	487	6	AX140716	AX140716 Sequence
C 31	439.6	37.4	487	6	AX200576	AX200576 Sequence
C 32	439.6	37.4	487	6	AX267232	AX267232 Sequence
C 33	439.6	37.4	487	6	BD070348	BD070348 Sequence
34	426.4	36.3	623	9	HSU44105	U44105 Human Raby
35	385.4	32.8	399	6	BD054210	BD054210 Sequence
36	333.2	28.4	1049	9	AB036693	AB036693 Homo sapi
C 37	333.2	28.4	31483	9	AL139228	AL139228 Human DNA
C 38	333.2	28.4	171430	2	AC027160	AC027160 Homo sapi
C 39	333	28.3	199658	2	AC133053	AC133053 Rattus no
C 40	329.2	28.0	212953	2	AC144376	AC144376 Pan trogl
C 41	328.4	27.9	160571	10	AL672008	AL672008 Mouse DNA
C 42	328.2	27.9	163495	9	AC024183	AC024183 Homo sapi
C 43	328.2	27.9	164891	2	AC022848	AC022848 Homo sapi
44	328.2	27.9	199174	9	AC007742	AC007742 Homo sapi
C 45	324.8	27.6	158422	9	AC026616	AC026616 Homo sapi

ALIGNMENTS

RESULT 1  
AR034083  
LOCUS AR034083 1175 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 6 from patent US 5869291.  
ACCESSION AR034083  
VERSION AR034083.1 GI:5949688  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1175)  
AUTHORS Hillman,J.L., Goli,S.K. and Bandman,O.  
TITLE Rab proteins  
JOURNAL Patent: US 5869291-A 6 09-FEB-1999;  
FEATURES Location/Qualifiers

```

source          1. 1175
BASE COUNT      387 a 191 c 244 g 353 t
ORIGIN

Query Match      100.0%; Score 1175; DB 6; Length 1175;
Best Local Similarity 100.0%; Pred. No. 3.5e-223;
Matches 1175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACTGTGATGAAACACATTTCCCGTGTCCGTTGAGTGCATCTTCTCAACACCTAGGAG 60
DB 1 CACTGTGATGAAACACATTTCCCGTGTCCGTTGAGTGCATCTTCTCAACACCTAGGAG 60

QY 61 GGTTCCTTGAAGCTTTTTCAGATTAACAATGGCAGGAAATCATCACATTTTAAAGTAATTC 120
DB 61 GGTTCCTTGAAGCTTTTTCAGATTAACAATGGCAGGAAATCATCACATTTTAAAGTAATTC 120

QY 121 TCCTTGAGATGTTGGAGTTGGGAAGATTCACATTAATGAACAGATATGTAACATAAGT 180
DB 121 TCCTTGAGATGTTGGAGTTGGGAAGATTCACATTAATGAACAGATATGTAACATAAGT 180

QY 181 TTGATACCCAGCTCTTCCATACATAGTGTGGAAATTTTAAATAAGATTTGGAAGTGG 240
DB 181 TTGATACCCAGCTCTTCCATACATAGTGTGGAAATTTTAAATAAGATTTGGAAGTGG 240

QY 241 ATGACATTTTGTACCATCAGATTTGGACACGCGCAGGTTCAGGAGCGATTCGGAAGCC 300
DB 241 ATGACATTTTGTACCATCAGATTTGGACACGCGCAGGTTCAGGAGCGATTCGGAAGCC 300

QY 301 TGAGGACACCATTTTACAGAGTTCTGACGTCTGCTCTTACTTTTAGTTCGATGAT 360
DB 301 TGAGGACACCATTTTACAGAGTTCTGACGTCTGCTCTTACTTTTAGTTCGATGAT 360

QY 361 CACAAAGCTTCCAGAACTTAAGTAACCTGGAAGAAAGAAATTCATATATTATGAGATGCA 420
DB 361 CACAAAGCTTCCAGAACTTAAGTAACCTGGAAGAAAGAAATTCATATATTATGAGATGCA 420

QY 421 AAGAGCTGAGAGCTTTCTTTTGTGATCTTGGTAAACAGATGTCATGACAGTCAATCTTACC 480
DB 421 AAGAGCTGAGAGCTTTCTTTTGTGATCTTGGTAAACAGATGTCATGACAGTCAATCTTACC 480

QY 481 AGGTGTCTACAGAAAGCCCAAGCTTGTGAGGACCAACGCGACTATCTTATTTTG 540
DB 481 AGGTGTCTACAGAAAGCCCAAGCTTGTGAGGACCAACGCGACTATCTTATTTTG 540

QY 541 AAACAGTGCAGAAAGATGCAAAATGTGCGCAGAGCTTTGAGGAAGCGGTTCGAAAG 600
DB 541 AAACAGTGCAGAAAGATGCAAAATGTGCGCAGAGCTTTGAGGAAGCGGTTCGAAAG 600

QY 601 TTCTTGCTACCGAGGATAGTGCAGATCATTTGATTCAGACACACAGTCAATCTTACC 660
DB 601 TTCTTGCTACCGAGGATAGTGCAGATCATTTGATTCAGACACACAGTCAATCTTACC 660

QY 661 GAAAGCCCAAGCTAGCTCATCTGCTGTGATGTTAGATTTGTTGATGCAATCTTAAACCA 720
DB 661 GAAAGCCCAAGCTAGCTCATCTGCTGTGATGTTAGATTTGTTGATGCAATCTTAAACCA 720

QY 721 ACTCACATATACACAAATCAACATGGGATGGAGAGAAATAGCGTTTCAGAG 780
DB 721 ACTCACATATACACAAATCAACATGGGATGGAGAGAAATAGCGTTTCAGAG 780

QY 781 TGTATCATCTACTATAAATTAACATAATGTTGCTGCTTCAATAGTTGGTGGGAGAAGG 840
DB 781 TGTATCATCTACTATAAATTAACATAATGTTGCTGCTTCAATAGTTGGTGGGAGAAGG 840

QY 841 GACATATCCTCTGGAGGATATATTACTCAATATGCAATGGCACTTCAATTAATTAAT 900
DB 841 GACATATCCTCTGGAGGATATATTACTCAATATGCAATATGGCACTTCAATTAATTAAT 900

QY 901 GTAACAGTTGTCTAATAAGCTTTCTTTAATTTAAATATGTAAGTTGCAAGCTAATAAT 960
DB 901 GTAACAGTTGTCTAATAAGCTTTCTTTAATTTAAATATGTAAGTTGCAAGCTAATAAT 960

```

---

```

QY 961 GAAATGACCAAGACTTTAATTATATAATAAAAAAAGAAACTTGACTATTCTAGAAATTATA 1020
DB 961 GAAATGACCAAGACTTTAATTATATAATAAAAAAAGAAACTTGACTATTCTAGAAATTATA 1020

QY 1021 CTTGATATTTTCTCGGAAATGAGAACTGAGAACTACTTTTATATGTTGTTATGTCGAA 1080
DB 1021 CTTGATATTTTCTCGGAAATGAGAACTGAGAACTACTTTTATATGTTGTTATGTCGAA 1080

QY 1081 TTAGCATTTGATTTCTGTTCCAGGAAATATCTTTCTTAAAGCAATATGTTAGATATAA 1140
DB 1081 TTAGCATTTGATTTCTGTTCCAGGAAATATCTTTCTTAAAGCAATATGTTAGATATAA 1140

QY 1141 AGATTTAAATCTAATGTTAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1175
DB 1141 AGATTTAAATCTAATGTTAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1175

RESULT 2
AC003037      213863 bp.   DNA      linear      PRI 30-MAR-2001
LOCUS      Homo sapiens Xp22 BAC GSHB 526D21 (Genome Systems Human BAC
DEFINITION      AC003037
ACCESSION      AC003037
VERSION      AC003037.1 GI:2920805
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 213863)
AUTHORS      Muhyi,D., Aronson,A.D., Brundage,E., Carvelli,K., Chen,J.,
Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R.,
Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L.,
Jin,S., Kampal,R., Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O.,
Lu,J., Ly,T., Martinez,C., Osval,G., Perez,L., Rashid,N.D.,
Rowland,K., Savage,L., Scherer,S.S., Shen,H., Tims,K.M., Todd,J.,
Vo,Q., Worley,K.C., Yu,W., Chinalt,C., Nelson,D. and Gibbs,R.A.
Direct Submission
Unpublished
REFERENCE      2 (bases 1 to 213863)
AUTHORS      Worley,K.C.
TITLE      Direct Submission
JOURNAL      Submitted (30-OCT-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE      3 (bases 1 to 213863)
AUTHORS      Worley,K.C.
TITLE      Direct Submission
JOURNAL      Submitted (28-FEB-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE      4 (bases 1 to 213863)
AUTHORS      Worley,K.C.
TITLE      Direct Submission
JOURNAL      Submitted (06-MAR-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE      5 (bases 1 to 213863)
AUTHORS      Worley,K.C.
TITLE      Direct Submission
JOURNAL      Submitted (01-OCT-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE      6 (bases 1 to 213863)
AUTHORS      Worley,K.C.
TITLE      Direct Submission
JOURNAL      Submitted (06-OCT-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE      7 (bases 1 to 213863)
AUTHORS      Worley,K.C.
TITLE      Direct Submission
JOURNAL      Submitted (30-MAR-2001) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Mar 2, 1998 this sequence version replaced gi:2804324.
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no

```

ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

## FEATURES

## source

```
1..213863
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xp22"
/clone="GSHB-526D21"
/clone_lib="Genome Systems Human BAC library"
/clone="1..1999"
```

```
misc_feature
/note="Overlaps bases 149277..151276 of clone AC079383"
/functions="Overlaps with adjacent clone AC079383"
```

```
misc_feature
1..303
```

```
misc_feature
/note="Region: AA029104 zk09b01.r1 Soares pregnant uterus complement(380..697)"
```

```
repeat_region
/note="Region: AA043674 zk50a09.s1 Soares pregnant uterus complement(1102..1291)"
```

```
repeat_region
```

```
repeat_region
```

```
misc_feature
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
misc_feature
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
repeat_region
```

```
15421..15925
/note="Region: AA828182 of33912.s1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1422982"
complement(16053..16074)
/rpt_family="AT-rich"
16136..16293
/rpt_family="AluSp"
16342..16644
/rpt_family="AluSx"
16710..16999
/rpt_family="AluSg"
17038..17335
/rpt_family="AluSx"
18365..18654
/rpt_family="AluSx"
20209..20243
/rpt_family="POD_Y_A"
20634..20920
/rpt_family="AluSx"
20946..21375
/rpt_family="MLT1C"
22045..22349
/rpt_family="AluSx"
23918..24220
/rpt_family="AluJb"
24370..24596
/rpt_family="MIR"
complement(24994..25130)
/rpt_family="MER58A"
25383..25414
/rpt_family="POLY_A"
25856..25953
/rpt_family="MER3"
25956..26288
/rpt_family="AluY"
complement(26289..26373)
/rpt_family="MER33"
complement(26397..26641)
/rpt_family="AluSx"
26965..27292
/rpt_family="MLT1D"
complement(27563..27593)
/rpt_family="AT-rich"
27807..32800
/rpt_family="L1PA2"
32380..32557
/standard_name="STS1-CSRL-31b6-uA/CSRL-31b6-uZ, Chr. X, Homo sapiens"
Homo sapiens
/db_xref="dbSTS:7821"
32801..33102
/rpt_family="AluSp"
33106..34245
/rpt_family="L1PA2"
34748..34864
/rpt_family="L2"
35687..35752
/standard_name="HUMSWX1791, Chr. X, Homo sapiens"
/db_xref="dbSTS:13087"
37266..37472
/rpt_family="MER20"
complement(40152..40194)
/rpt_family="L2"
complement(40214..40291)
/rpt_family="MIR"
complement(40560..40761)
/rpt_family="L1MC1"
40763..40844
/rpt_family="AluSg/x"
complement(41023..41322)
/rpt_family="AluSx"
complement(42296..42322)
/rpt_family="AT-rich"
42638..42936
```

Query Match		93.6%	Score 1100;	DB 9;	Length 213863;
Best Local Similarity		100.0%	Pred. No. 9.3e-209;		
Matches 1100;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	59	AGGTTCTTGAAGCTTTGAGATTAAACATAGGTGGAGAAATCATCTTTTAAAGTAAT	118		
Db	11893	AGGTTCTTGAAGCTTTGAGATTAAACATAGGTGGAGAAATCATCTTTTAAAGTAAT	11952		
Qy	119	TCCTCTGGAGATGGTGGAGTTGGAGAGTTCACTATGACAGATATGTAACHTAA	178		
Db	11953	TCCTCTGGAGATGGTGGAGTTGGAGAGTTCACTATGACAGATATGTAACHTAA	12012		
Qy	179	GTTCGATACCCAGCTTTCATCAATAGGTGGAGTTTAAATAAGAGATTGGAAGT	238		
Db	12013	GTTCGATACCCAGCTTTCATCAATAGGTGGAGTTTAAATAAGAGATTGGAAGT	12072		
Qy	239	GGATGGACATTTTACCAATGAGATTGGGACAGCGAGGTGAGGAGCGATTCGAG	298		
Db	12073	GGATGGACATTTTACCAATGAGATTGGGACAGCGAGGTGAGGAGCGATTCGAG	12132		
Qy	299	CCTGAGGACACCATTTTACAGAGTTCTGACTGCTGCTTACCTTTTAGTGTGATGA	358		
Db	12133	CCTGAGGACACCATTTTACAGAGTTCTGACTGCTGCTTACCTTTTAGTGTGATGA	12192		
Qy	359	TTCAAAAGCTTCAGAACTTAAGTAACTGGAGAAAGAAATCATATATATGCGATGT	418		
Db	12193	TTCAAAAGCTTCAGAACTTAAGTAACTGGAGAAAGAAATCATATATATGCGATGT	12252		
Qy	419	GAAGAGCCTGAGAGCTTTCCTTTTCTGATTTCTGGTTAAACAGATTGACATAAGCGAAG	478		
Db	12253	GAAGAGCCTGAGAGCTTTCCTTTTCTGATTTCTGGTTAAACAGATTGACATAAGCGAAG	12312		
Qy	479	CGAGTGTCTACAGAAAGCCCAAGCTTGTGTCAGGACCAACGCGACTATCTTATT	538		
Db	12313	CGAGTGTCTACAGAAAGCCCAAGCTTGTGTCAGGACCAACGCGACTATCTTATT	12372		
Qy	539	TGAACAGAGTGCAGAAAGATGCCAATGTGCGAGCAGCTTTGAGGAGCGTTTCGAG	598		
Db	12373	TGAACAGAGTGCAGAAAGATGCCAATGTGCGAGCAGCTTTGAGGAGCGTTTCGAG	12432		
Qy	599	AGTCTTGTCTACAGAGATAGTTCAGATCATTTGATTCAGACAGACACAGTCAATCTCA	658		
Db	12433	AGTCTTGTCTACAGAGATAGTTCAGATCATTTGATTCAGACAGACACAGTCAATCTCA	12492		
Qy	659	CGAAAGCCCAAGCCTAGCTCATCTTCTGTTGATTTGATTTGATTTGATTTGATTTGAT	718		
Db	12493	CGAAAGCCCAAGCCTAGCTCATCTTCTGTTGATTTGATTTGATTTGATTTGATTTGAT	12552		
Qy	719	CAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGATTAGCGTTTGCAGC	778		
Db	12553	CAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGATTAGCGTTTGCAGC	12612		
Qy	779	AGTGTATCATCTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	838		
Db	12613	AGTGTATCATCTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	12672		
Qy	839	GGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCCACTTTACATTTATATA	898		
Db	12673	GGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCCACTTTACATTTATATA	12732		
Qy	899	TTGTAACAGTGTCTAATAACGTTCTTTTAAATTTAAATGTAAGTTCAGAGCTTAATA	958		
Db	12733	TTGTAACAGTGTCTAATAACGTTCTTTTAAATTTAAATGTAAGTTCAGAGCTTAATA	12792		
Qy	959	ATGAATGACCAAGACTTTTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1018		
Db	12793	ATGAATGACCAAGACTTTTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA	12852		
Qy	1019	TACTTGGATTTTCTCGGAAATAGGAACTACTTTTTATATGATGTATGTTTTATGTC	1078		
Db	12853	TACTTGGATTTTCTCGGAAATAGGAACTACTTTTTATATGATGTATGTTTTATGTC	12912		

Qy	1079	AATTAGCATTCGATTCCTTGGTTTCAGGGAATACCTTCTCTAAGCAATAATGTTAGATATT	1138		
Db	12913	AATTAGCATTCGATTCCTTGGTTTCAGGGAATACCTTCTCTAAGCAATAATGTTAGATATT	12972		
Qy	1139	AAAGATTAAATCTAATGTA	1158		
Db	12973	AAAGATTAAATCTAATGTA	12992		
RESULT 3					
BC017265		1106 bp	linear	PRI 09-NOV-2001	
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REMARK					
COMMENT					
<p>NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a></p> <p>Contact: MGC help desk</p> <p>Email: <a href="mailto:cgaps-xemail.nih.gov">cgaps-xemail.nih.gov</a></p> <p>Tissue Procurement: ATCC</p> <p>cDNA Library Preparation: Rubin Laboratory</p> <p>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)</p> <p>DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;</p> <p>Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a></p> <p>Contact: <a href="mailto:nisc.mgc@hgr.nih.gov">nisc.mgc@hgr.nih.gov</a></p> <p>Shenchenko, X., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blackley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karling, E., Legaspi, R., Lim, M., Maduro, Q.L., Masello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tongson, E.B., Touchman, J.W., Tsourgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.</p>					
<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a></p> <p>Series: IRAL Plate: 39 Row: c Column: 18.</p>					
<p>FEATURES</p> <p>Location/Qualifiers</p> <p>1..1106</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="LocusID:9367"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="MGC:29629 IMAGE:4139714"</p> <p>/tissue_type="Muscle, zhabdomyosarcoma"</p> <p>/clone_lib="NIH MGC 17"</p> <p>/lab_host="DH10B-R"</p> <p>/notes="Vector: pOTB7"</p> <p>192..797</p> <p>/codon_start=1</p> <p>/product="RAB9, member RAS oncogene family"</p> <p>/protein_id="AAH17265.1"</p> <p>/db_xref="GI:16878118"</p> <p>/translation="MAGKSLFKVILLGGVGKSLMRYVTKNFKDTQLFHTIGVEF LNKEDVDGHFTVQIWDVAGGFRSLFTPFYRSDCCLLTFVSDSQQFNLNWK KEFYIYADVKPEPFPFVLGNKIDISERQVSTERAQWCRDNGDYPYETSAXDATN VAAAPFEAVRRVLATEDRSDHLIQTDTVNLHKKPKPSSCC"</p>					
<p>CDS</p> <p>339 a 210 c 249 g 308 t</p>					
<p>BASE COUNT</p> <p>ORIGIN</p>					

Query Match		84.5%;	Score 993.2;	DB 9;	Length 1106;
Best Local Similarity		99.7%;	Pred. No. 4.4e-187;		
Matches 995;		Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;
QY	1	CACCTGTGATGAACACCTTTCCCGTGTGCTTTGAGTGCATCTTCTCAACACCCCTAGGAG	60		
Db	106	CACCTGTGATGAACACCTTTCCCGTGTGCTTTGAGTGCATCTTCTCAACACCCCTAGGAG	165		
QY	61	GCTTCTTCAAGCTTTTGAGATTAAACAATGGCAGGAATCATCACTTTTAAAGTAATTC	120		
Db	166	GCTTCTTCAAGCTTTTGAGATTAAACAATGGCAGGAATCATCACTTTTAAAGTAATTC	225		
QY	121	TCCTTGGAGATGGTGGAGTTGGGAAGCTTCACTTATGAACAGATATGTAATAAAGT	180		
Db	226	TCCTTGGAGATGGTGGAGTTGGGAAGCTTCACTTATGAACAGATATGTAATAAAGT	285		
QY	191	TTGATACCCAGCTCTTCCATACATAGTGTGGATTTTAAATAAAGATTTGGAAGTGG	240		
Db	286	TTGATACCCAGCTCTTCCATACATAGTGTGGATTTTAAATAAAGATTTGGAAGTGG	345		
QY	241	ATGGACATTTTGTACCATGACAGTTTGGGACACGGCAGGTCCAGAGCGATTCGGAAGCC	300		
Db	346	ATGGACATTTTGTACCATGACAGTTTGGGACACGGCAGGTCCAGAGCGATTCGGAAGCC	405		
QY	301	TGAGGACACCATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTGTGTCGATGATT	360		
Db	406	TGAGGACACCATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTGTGTCGATGATT	465		
QY	361	CACAAAGCTTCCAGAACTTAACTAGTAACCTGGAAGAAAGTAATCATATATTATGACAGATGTA	420		
Db	466	CACAAAGCTTCCAGAACTTAACTAGTAACCTGGAAGAAAGTAATCATATATTATGACAGATGTA	525		
QY	421	AAGAGCTGTGAGCTTTTCTTTTGTGATTTCTGGGTAAACAAGTTGACATAAGCGAACCGC	480		
Db	526	AAGAGCTGTGAGCTTTTCTTTTGTGATTTCTGGGTAAACAAGTTGACATAAGCGAACCGC	585		
QY	481	AGGTGTCTACAGAAAGACCCCAAGCTTGTGTCAGGACCAACGCGACTATCTTATTTTG	540		
Db	586	AGGTGTCTACAGAAAGACCCCAAGCTTGTGTCAGGACCAACGCGACTATCTTATTTTG	645		
QY	541	AAACAAGTGCAAAAGATGCGCAAAATGTGCGACAGCCTTTGAGGAAGCGGTTTCGAAGAG	600		
Db	646	AAACAAGTGCAAAAGATGCGCAAAATGTGCGACAGCCTTTGAGGAAGCGGTTTCGAAGAG	705		
QY	601	TTCTTGCTACCGAGGATAGTGCATGATTTGATTCAGACAGACAGTCATCTTCACC	660		
Db	706	TTCTTGCTACCGAGGATAGTGCATGATTTGATTCAGACAGACAGTCATCTTCACC	765		
QY	661	GAAGGCCAAGCCTAGCTCATCTTGTGTTGATTGTTAGATTGTTGATGCAATCTAACA	720		
Db	766	GAAGGCCAAGCCTAGCTCATCTTGTGTTGATTGTTAGATTGTTGATGCAATCTAACA	825		
QY	721	ACTCAGACATATACAGAAATCAACATGGGGATGGGAAGAGAAATAGCGTTTGCAGCAG	780		
Db	826	ACTCAGACATATACAGAAATCAACATGGGGATGGGAAGAGAAATAGCGTTTGCAGCAG	885		
QY	781	TGTATCATCTPACTAATAAATAAATACTAATGTTGCTGCTTCACTTGTGTTGGGAGAG	840		
Db	886	TGTATCATCTPACTAATAAATAAATACTAATGTTGCTGCTTCACTTGTGTTGGGAGAG	945		
QY	841	GACACATCCACTCTTGGAGAAATATATTTACTCAATAATGGCAACCTTACATTTATAAAT	900		
Db	946	GACACATCCACTCTTGGAGAAATATATTTACTCAATAATGGCAACCTTACATTTATAAAT	1005		
QY	901	GTAACAGTTGTCTAATAACCTTTCTTTTAAATATGATGTTGTCAGAGCTAATAAT	960		
Db	1006	GTAACAGTTGTCTAATAACCTTTCTTTTAAATATGATGTTGTCAGAGCTAATAAT	1065		
QY	961	GAATGACCAAGACTTAAATATAATAAATAAAGAA	998		
Db	1066	GAATGACCAAGACTTAAATATAATAAATAAAGAA	1103		

RESULT 4

AC091987/c

LOCUS

DEFINITION

AC091987

AC091987

VERSION

AC091987.2

GT:20376969

HTG.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 143872)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Unpublished

2 (bases 1 to 143872)

DOE Joint Genome Institute.

Direct Submission

Submitted (09-JUN-2001)

Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 143872)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Submitted (01-MAY-2002)

DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On May 1, 2002 this sequence version replaced gi:14333923.

Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu

Quality: Phrap Quality >=40 99.4% of Sequence;

Estimated Total Number of Errors is 0.6.

NOTE: This insert is not the entire sequence of the clone (entire sequence is 149.5kb). It is clipped at the overlap with AC010590.

The number of bases overlapped is 20375.

location/Qualifiers

1..143872

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="RP11-6E2"

BASE COUNT

45872 a 25416 c 25617 g 46967 t

ORIGIN

Query Match

Best Local Similarity

Matches 970;

Conservative 0;

Mismatches 113;

Indels 37;

Gaps 5;

QY

86

AATGGCAGGAAATCATCACTTTTAAAGTAATTTCTCTTGGAGATGGTGGAGTTGGAA

145

Db

26969

AACGGCAGGAAATCATCACTTTTAAAGTAATTTCTCTTGGAGATGGTGGAGTTGGAA

26910

QY

146

GAGTTCACTTATGAACAGATATGAATTAAGTTTGAATGATACCCAGCTTCCATCAAT

205

Db

26909

GAGTTCACTTATGAACAGATATGAATTAAGTTTGAATGATACCCAGCTTCCATCAAT

26850

QY

206

AGGTGTGGAATTTTAAATTAAGATTTGGAATGGAATGGAATTTGGTACCAAGTAGAA

265

Db

26849

AGGTGTGGAATTTTAAAGTAAGATTTGGAATGGAATGGAATTTGGTACCAAGTAGAA

26790

QY

266

TTGGGACACGCGAGTCCAGGAGGATTCGGAAGCTGAGACACCATTTTACAGAGTTTC

325

Db

26789

TTGGGACACGCGAGTCCAGGAGGATTCGGAAGCTGAGACACCATTTTACAGAGTTTC

26730

QY

326

TGACTGCTGCTGCTTACTTTTGTAGTGTGATGATTCACAAAGCTTCCAGAACTTAAGTAA

385

Db

26729

TGACTGCTGCTGCTTACTTTTGTAGTGTGATGATTCACAAAGCTTCCAGAACTTAAGTAA

26671

QY

386

CTGGAAGAGAAATTCATATATTATGAGATGTGAAGAGCTTCAGAGCTTCTCTTTTGT

445

Db

26670

CTGGAAGAGAAATTCATATATTATGAGATGTGAAGAGCTTCAGAGCTTCTCTTTTGT

26611

```

QY 446 GATTCTGGTAAACAAGATTGACATAGCGAAGCGAGGTGTCTACAGAGAAGCCCAAGC 505
Db 26610 AATTTTGGGTAAACAAGCTGAC---AGTGAACAGCAGGTGTCTGCAAGAGAAGCCCAAGC 26554
QY 506 TTGGTGAGGAGCAACGGCGACTATCTCTTATTTTGAACAAGTGCACAAAGATGCCACAA 565
Db 26553 CTGGTGAGGAGCAACGGCGACTATCTCTTATTTTGAACAAGTGCACAAAGATGCCACAA 26494
QY 566 TGTGGCAGCAGCCTTTTGAAGAGCGGTTCGAAGAGTCTTGTACCGAGATAGTTCAGA 625
Db 26493 TGTGGCAGTACGATTTTGAAGAGCGGTTCGAAGAGTCTTGTACCGAGATAGTTCAGA 26434
QY 626 TCATTGATTTCAGACAGACA-----CACTCAATCTTCCACCA 662
Db 26433 TCATTGACTCAGAGAGTGTGTCTGACTTCATCGAAGAACCATTTGTCAATCTTCATTGA 26374
QY 663 AAGCCCAAGCCTAGCTCAGCTCTCTCTGTGATGTT-----AGATTGTTGATGATCTTAAC 718
Db 26373 AAGCCCAAGCCTAGCTCAGCTCTCTCTGTGATGTT-----AGATTGTTGATGATCTTAAC 26314
QY 719 CAACTCACACATATACACAAATCAACATGGGGATGGAGAAGAGATTAGCGTTTGCAGC 778
Db 26313 CAACTCACAGGTACACACAAATCAACATGGGGATGGAGAAGAGATTAGCGTTTGCAGC 26254
QY 779 AGTGATCATCTACTAATAAATAAATAAATAATGT-----TGCTGCTTCATTAGTTGGTG 832
Db 26253 AGTGATCATCTACTAATAAATAAATAAATAATGT-----TGCTGCTTCATTAGTTGGTG 26194
QY 833 GGAGAAGGAGCACATCCACTCTTGGAGGAATATATTTACTCAATATGACCACTTACATT 892
Db 26193 GGAGAAGGAGCACATCCACTCTTGGAGGAATATATTTACTCAATATGACCACTTACATT 26134
QY 893 TATAATTGTAACAGTGTCTAATAACGTTCTTAAATTAATTAATGTAAGTTGCAGAGC 952
Db 26133 TATAATTGTAACAGTGTCTAATAACGTTCTTAAATTAATTAATGTAAGTTGCAGAGC 26074
QY 953 TAATAAATGAATGACCAAGACTTTAATATATAATAAATAAATAAATAAATAAATAAATAA 1012
Db 26073 TAATAAATGAATGACCAAGACTTTAATATATAATAAATAAATAAATAAATAAATAAATAA 26014
QY 1013 AGTTATATCTGGATTTTCTCGGAAATATGGAAGACTTCTTTTATATGATGATGTT 1072
Db 26013 AAATTATGCTGGATTTTCTCGGAAATATGGAAGACTTCTTTTATATGATGATGTT 25954
QY 1073 TTATGCAATTAGCATTTGATTTCTGTTCCAGGAATACTTTCTTAAGCAATATGTA 1132
Db 25953 TTATGCAATTAGCATTTGATTTCTGTTCCAGGAATACTTTCTTAAGCAATATGTA 25894
QY 1133 GATATTAAAGATTAATAATCTAATGTAAATAAATAAATAAATAAATAAATAAATAAATAA 1172
Db 25893 GATATTAAAGATTAATAATCTAATGTATTTGAAATAAATAAATAAATAAATAAATAAATAA 25854

RESULT 5
CFRAB9
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CFRAB9
Canine rab9 mRNA for ras-related GTP-binding protein.
X56386
GI:2189
GTP-binding protein; protein transport; ras gene; ras gene family.
Canis familiaris (dog)
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1
Chavrier, P., Vingron, M., Sander, C., Simons, K. and Zerial, M.
Molecular cloning of YP1/SEC4-related cDNAs from an epithelial
cell line
Mol. Cell. Biol. 10 (12), 6578-6585 (1990)
PUBMED
2123294
2 (bases 1 to 1001)
REFERENCE
AUTHORS
Lombardi, D., Soldati, T., Riederer, M.A., Goda, Y., Zerial, M. and

```

```

Pfeffer, S.R.
Rab9 functions in transport between late endosomes and the trans
Golgi network
EMBO J. 12 (2), 677-682 (1993)
93178443
PUBMED
8440258
3 (bases 1 to 1001)
Zerial, M.
Direct Submission
Submitted (17-AUG-1990) Zerial M., EMBL, Meyerhofstrasse 1, 6900
Heidelberg, FRG
FEATURES
source
1..1001
/organism="Canis familiaris"
/mol_type="mRNA"
/strain="cocker spaniel"
/db_xref="taxon:9615"
/clone="15"
/cell_line="MDCK"
1..1001
/gene="rab9"
<1..519
/gene="rab9"
<1..519
/gene="rab9"
/codon_start=1
/protein_id="CAA39797.1"
/db_xref="GI:2190"
/db_xref="SWISS-PROT:P24408"
/translations="NKFDTQLPHTIGVFLNKLEVDGHPVTMQIWDVTAGQRRFSLR
TPYRSDCLLUTFSVDDSQNSUNWKEFIYADVKEPSPFVILGNKIDISER
QVSTESAQAWCNDNGDYPYFTSAKDATNVAAFEEAVRRLVATSDRLIQDTVS
LHKPKPSSSC"
<1..516
/gene="rab9"
/products="rab9 GTP-binding protein"
BASE COUNT 338 a 174 c 210 g 279 t
ORIGIN
Query Match 60.7%; Score 713.6; DB 4; Length 1001;
Best Local Similarity 87.1%; Pred. No. 1.4e-131;
Matches 881; Conservative 0; Mismatches 109; Indels 22; Gaps 8;
QY 174 AATAAGTTTGATACCCAGCTCTTCCATCAATAGGTGTGGAATTTTAAATAAAGATTG 233
Db 1 AATAAGTTTGATACCCAGCTCTTCCATCAATAGGTGTGGAATTTTAAATAAAGATTG 60
QY 234 GAAGTGGATGGACATTTTGTACCATGCAGATTGGGACACGCGAGGTCCAGGCGATTTC 293
Db 61 GAGTGGATGGACATTTTGTACCATGCAGATTGGGACACGCGGTCCAGGCGATTTC 120
QY 294 CGAAGCTGAGGACACCATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAGTGTC 353
Db 121 AGAAGCTGAGGACGCGCTTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAGTGTC 180
QY 354 GATGATTTCAGAGCTTCAGAACTTAAGTAAGTGAAGAAAGAAATTCATATATTATGCA 413
Db 181 GATGATTTCAGAGCTTCAGAACTTAAGTAAGTGAAGAAAGAAATTCATATATTATGCA 240
QY 414 GATGTAAGAGAGCTGAGAGCTTTCTTTTGTGATTTCTGGGTAAACAAGATTGACATAAGC 473
Db 241 GATGTAAGAGAGCGCCGAAAGCTTTCTTTTGTGATTTTGGGCAACAAGATCGACATAAGT 300
QY 474 GAAGCGAGGTGTCTACAGAAGAGCCCAAGCTTGGTCAGGGAACAAGCGGACTATCTCT 533
Db 301 GAAGCGAAGTGTCTACAGAAGAGCCCAAGCTTGGTCAGGGAACAAGCGGACTATCTCT 360
QY 534 TATTTTGAACAAAGTGCACAAAGATGCCAATAATGTGGCAGAGCGCTTTTGAGGAAGCGTT 593
Db 361 TACTTTGAACAAAGTGCACAAAGATGCCAATAATGTGGCAGAGCGCTTTTGAGGAAGCGTT 420
QY 594 CGAAGATTTCTGCTACCGAGGATAGGTGATGATTTGATTCACACAGACAGTCAAT 653

```

Db 421 CGAAGAGTGTCTACTAGGATAGGTGACATGACCTGATTCAGACAGACACAGTCAGC 480

Qy 654 CTTACCGAAGCCAGCCTAGCTCATCTTGTCTGTGATGTTAGA----TTGTTGATG 709

Db 481 CTGACCGAAGCCAGCCTAGCTCATCTTGTCTGTGATGTTAGA----TTGTTGATG 540

Qy 710 CATTTCAACCAATCTCACATATATACAAAATCAACATGGGGATGGAGAGAGAAATTAGC 769

Db 541 CGATCTAACCAATCTCACATATATACAAAATCAACATGGGGATGGAGAGAGAAATTAGC 599

Qy 770 GTTTCGAGCAGTATCATCTCATATAAAATTAACCTA---ATGTTGCTGCTTCATTAG 826

Db 600 ATTTCGAGCAATGATCCTACTCATATAAAATTAACCTA---ATGTTGCTGCTTCATTAG 659

Qy 827 TTGTTGGGAGAGGACACATCACTC---TTGGAGGAATATATTTACTCAATATGACACC 885

Db 660 TGGGTTGGGAGAGGACACATCACTC---TTGGAGGAATATATTTACTCAATATGACACC 719

Qy 886 TTACATTTATAATTTGTAACAGTGTCTATAAATACGTTCTTTTAATTTAAATATGTAAGTT 945

Db 720 TTACATTTATAATTTGTAATGTTGTCTATAACG---TTTAATTTAAATATGTAAGTT 775

Qy 946 GCAGAGCTAATAATGAATGACCAAGCTTTAATTAATAAATAAATAAGAACTTGAAT 1005

Db 776 ACAGAGCT-ATAACGAGATGATCAAGACTTTTAATTAACAT-----TAAACACTTGAAT 829

Qy 1006 ATTCTAGAAGTTATCTGATTTCTCTGGAAATGAGAACTACTTTTATATATG 1065

Db 830 ATTCTAGAAGTTAATGCTTTTCTCTGGAAATGAGAACTACTTTTATATATG 889

Qy 1066 TATGTTTATGCAATGATGATTTCTGTTCTCA---GGGAATATCTTCTTAAGC 1122

Db 890 TATATTTTGTGTAATGATGATTTCTGTTCTGAGGGGGAAGATTCCTTAAGC 949

Qy 1123 AATAATGTTAGATTAATAAGATTAATAATCTAAATGTAATAAAAAAAAAAAAA 1174

Db 950 AATAATGTTAATAATAAGATTAATAATCTAAATAAAAAAAAAAAAA 1001

RESULT 6

AC084354/c

LOCUS

DEFINITION

AC084354 Homo sapiens chromosome RP11-11 clone RP11-809H9, WORKING DRAFT

SEQUENCE

AC084354.2 GI:13270842

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 130407)

Waterston,R.H.

The sequence of Homo sapiens clone

Unpublished

2 (bases 1 to 130407)

Waterston,R.H.

Direct Submission

Submitted (25-OCT-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA

On Mar 10, 2001 this sequence version replaced gi:10999052.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H NH0809H09

----- Summary Statistics -----

Sequencing vector: M13; 100%

Sequencing vector: plasmid; 0%

Chemistry: Dye-primer ET; 100% of reads

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 117221 bases at least Q40

Consensus quality: 121718 bases at least Q30

Consensus quality: 123447 bases at least Q20

Insert size: 193000; agarose-fp

Insert size: 121707; sum-of-contigs

Quality coverage: 2.39 in Q20 bases; agarose-fp

Quality coverage: 3.68 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 1423: contig of 1423 bp in length

\* 1424 1523: gap of unknown length

\* 1524 3282: contig of 1759 bp in length

\* 3283 3282: gap of unknown length

\* 3283 5435: contig of 2053 bp in length

\* 5436 5535: gap of unknown length

\* 5536 7716: contig of 2181 bp in length

\* 7717 7816: gap of unknown length

\* 7817 9788: contig of 1972 bp in length

\* 9789 9888: gap of unknown length

\* 9889 11982: contig of 2094 bp in length

\* 11983 12082: gap of unknown length

\* 12083 14414: contig of 2332 bp in length

\* 14415 14514: gap of unknown length

\* 14515 16518: contig of 2004 bp in length

\* 16519 16619: gap of unknown length

\* 16619 18797: contig of 2179 bp in length

\* 18798 18897: gap of unknown length

\* 18898 20913: contig of 2022 bp in length

\* 20920 21019: gap of unknown length

\* 21020 23200: contig of 2801 bp in length

\* 23201 23921: gap of unknown length

\* 23922 26571: contig of 2651 bp in length

\* 26572 26672: gap of unknown length

\* 26672 28477: contig of 1805 bp in length

\* 28477 28577: gap of unknown length

\* 28577 31317: contig of 2740 bp in length

\* 31317 33788: contig of 2372 bp in length

\* 33789 36379: gap of unknown length

\* 36380 36479: gap of unknown length

\* 36480 40039: contig of 3560 bp in length

\* 40040 40139: gap of unknown length

\* 40140 45356: contig of 5217 bp in length

\* 45357 45457: gap of unknown length

\* 45457 50509: contig of 5053 bp in length

\* 50510 50610: gap of unknown length

\* 50610 55122: contig of 4513 bp in length

\* 55123 55223: gap of unknown length

\* 55223 59639: contig of 4416 bp in length

\* 59640 64842: contig of 5104 bp in length

\* 64843 73926: gap of unknown length

\* 73927 74026: contig of 8984 bp in length

\* 74027 86109: contig of 12083 bp in length

\* 86110 86209: gap of unknown length

\* 86210 104005: contig of 17796 bp in length

\* 104006 127523: gap of unknown length

\* 127524 127624: gap of unknown length

\* 127625 128816: contig of 1193 bp in length

\* 128817 130407: contig of 1491 bp in length.





```

RESULT 7
AC022486
LOCUS       AC022486               156357 bp    DNA    linear    PRI 09-MAY-2001
DEFINITION  Homo sapiens BAC clone RP11-569J3 from Y, complete sequence.
ACCESSION   AC022486
VERSION     AC022486.4  GI:13435294
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156357)
AUTHORS     Sulston, J.E. and Waterston, R.
TOWARD      Toward a complete human genome sequence
JOURNAL     Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE     99063792
PUBMED      9847074
REFERENCE   2  (bases 1 to 156357)
AUTHORS     Scott, K., Maupin, R. and Swift, G.
TOWARD      The sequence of Homo sapiens BAC clone RP11-569J3
JOURNAL     Unpublished
REFERENCE   3  (bases 1 to 156357)
AUTHORS     Waterston, R.H.
TOWARD      Direct Submission
JOURNAL     Submitted (04-FEB-2000) Genome Sequencing Center, Washington
MO 63108, USA
REFERENCE   4  (bases 1 to 156357)
AUTHORS     Waterston, R.H.
TOWARD      Direct Submission
JOURNAL     Submitted (23-MAR-2001) Genome Sequencing Center, Washington
MO 63108, USA
REFERENCE   5  (bases 1 to 156357)
AUTHORS     Waterston, R.
TOWARD      Direct Submission
JOURNAL     Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
ON Mar 23, 2001 this sequence version replaced gi:7684584.
----- Genome Center
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0569J03
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The position of this clone was established as part of a
collaboration between the Human Chromosome Y Mapping Project
(Tomoko Kawaguchi, Helen Skalecky, Laura G. Brown, Steve Rozen,
and David C. Page at the Whitehead Institute for Biomedical
Research, Cambridge MA) and the Washington University Genome
Sequencing Center, St. Louis MO.

SOURCE INFORMATION:
The RP11-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,

```

Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bapac.med.buffalo.edu>)

VECTOR: pBAC3.6  
Location/Qualifiers

# NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-243P9, 200 base pair overlap; the clone sequenced to the right is RP11-143C1, 200 base pair overlap. Actual start of this clone is at base position 195 of RP11-569J3; actual end is at base position 14144 of RP11-143C1.

## FEATURES

source	
1..156357	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="Y"
	/map="Y"
	/clone="RP11-569J3"
	/clone_lib="RPCI-11"
9..1136	/rpt_family="Alu"
1726..2011	/rpt_family="Alu"
2630..2744	/rpt_family="Alu"
2745..2945	/rpt_family="Alu"
3157..3269	/rpt_family="Alu"
3780..4513	/rpt_family="L1"
4672..5268	/rpt_family="ERV1"
5274..5720	/rpt_family="L1"
5712..5907	/rpt_family="L1"
5911..6016	/rpt_family="L1"
6104..6398	/rpt_family="Alu"
6104..6173	/note="similar to EST AA864600 (NID:G2958913) od45a04.s1"
6601..6662	/rpt_family="L1"
6930..7409	/rpt_family="ERVK"
7686..7756	/rpt_family="ERV1"
11136..11427	/rpt_family="Alu"
12678..12976	/rpt_family="L1"
12980..14310	/rpt_family="L1"
14831..14990	/rpt_family="L1"
14991..15292	/rpt_family="Alu"
15293..15421	/rpt_family="L1"
15606..15804	/rpt_family="L1"
15882..15905	/rpt_family="L1"
15906..16178	/rpt_family="Alu"
16179..16385	/rpt_family="L1"
17438..18006	/rpt_family="ERV1"



```

SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE   1 (bases 1 to 183293)
AUTHORS    Wilson, R.K.
TITLE      The sequence of Pan troglodytes clone
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 183293)
AUTHORS    Wilson, R.K.
TITLE      Direct Submission
JOURNAL    Submitted (27-MAR-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT     ----- Genome Center -----
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@wustl.edu
            ----- Project Information -----
            Center project name: C P7005110
            ----- Location/Qualifiers -----
FEATURES             Location/Qualifiers
     source           1..183293
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="Y"
                     /clone="RP43-5110"
BASE COUNT          53050 a 34436 c 35146 g 60661 t
ORIGIN

Query Match      58.4%; Score 686.2; DB 9; Length 183293;
Best Local Similarity 81.5%; Pred. No. 1.3e-126;
Matches 909; Conservative 0; Mismatches 158; Indels 48; Gaps 8;

QY 59 AGGTTCTTCAAGCTTTTGAGATTACAAATGCGGAGGAATCATCTCTTTTAAAGTAAT 118
DB 31356 AGGTTCTTCAAGCTTTTGAGATTACAAATGCGGAGGAATCATCTCTTTTAAAGTAAT 31297

QY 119 TCTCTTGGAGATGGTGGAGTTGGGAGAGTTCACCTTATGAACAGATATGTAACTAATAA 178
DB 31296 TCTCTTGGAGATGGTGGAGTTGGGAGAGTTCACCTTATGAACAGATATGTAACTAATAA 31237

QY 179 GTTGTATACCACTCTTCCATACATAGTGTGGATTTTAAATAAGATTGGAAT 238
DB 31236 GTTGTGCACTCAGCTGTTCATCAATAGATGTGGAAATTTTAAATAA--GATTTGAAAG 31179

QY 239 GGATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTCAGGAGCGATTCGGAAG 298
DB 31178 GGATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTCAGGAGCGATTCGGAAG 31119

QY 299 CTTGAGACACATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTGTGTGATGCA 358
DB 31118 CTTGAGACACATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTGTGTGATGCA 31059

QY 359 TTCACAAAGCTTCAGAACTTAACTAGTAACTGGAAGAGATTCATATATATGCGAGTCT 418
DB 31058 TTTAGGAGCTTCAGAACTTAACTAGTAACTGGAAGAGATTCATATATATGCGAGTCT 41029

QY 419 GAAAGAGCTTCAGAGCTTTCTTTTGTGATCTCTGGTAAACAAGATTGACATAAGCGAAG 478
DB 31028 GAAAGAGCTTCAGAGCTTTCTTTTGTGATCTCTGGTAAACAAGATTGACATAAGCGAAG 30969

QY 479 GCAGGTGCTTACAGAGAGCCCAAGCTTGGTCAGGACCAAGCGGATATCTCTATTT 538
DB 30968 GCAGGTGCTTACAGAGAGCCCAAGCTTGGTCAGGACCAAGCGGATATCTCTATTT 30910

QY 539 TGAACAAGTGCAAAAGATGCCAACAAATGTGGCAGCAGCTTTTGAGGAAGCGGTTCCGAG 598
DB 30909 TGTAAACAAGTGCAAAAGATGCCAACAAATGTGGCAGCAGCTTTTGAGGAAGCGGTTCCAAA 30850

QY 599 AGTTCTTGCTACCGAGGATAGTTCAGATCATTTTGTATTCAGACAGACAGTCAATCTTCA 658

```

```

DB 30849 AATTTCTGGAGCTGGGATACGTACAGTCACTTGAATGAACACAGCATGTTCAACCTTCA 30790
QY 659 CGAAAGCCCAAGCCTAGCTCATCTTGTGTTGTTAGATT-----GTGTGATGCAATTC 714
DB 30789 CCAAAGCCCAAGCCTAGCAGATCTTGTATTGATGTTAGAGTTAAGCGAGATGCAATTC 30730
QY 715 TAACCAACTCACATATACACAAAATCAACATGCGGATGAGAGAGAAATAGCGTTTG 774
DB 30729 TAACCAATTCACATATACACAAAATCAACATGCGAGATGAGAGAAATAGCGTTTG 30670
QY 775 CAGCAGTGTATCATCTACTATAATAATAATAATAATAATAATAATAATAATAATAATAATA 830
DB 30669 TAGCAGTGTATCTACTATAATAATAATAATAATAATAATAATAATAATAATAATAATA 30610
QY 831 TGGGAGAGGAGACATCCACTCTTGGAGGAATATATTTACTCAATAATGCGACCTTACA 890
DB 30609 TGGGAGATGGGACATATCCACTCATGGAAGATATATTTCTCAATAATGAACCTTACA 30550
QY 891 TTTATAAATTTGACAGTTGCTTAATAC-----GTTTCTTTTAAATTAATATGTAAGTT 945
DB 30549 CTTATAAATTTGATGTTGTTGCTTAAACCAATGTTTCTTTGATTACATGCAATTA 30490
QY 946 GCAGAGCTAATAAATGAATGACCAAGACTTTAAATTAATAATAATAATAATAATAATAATA 1005
DB 30489 AGAGAGCTAATAAATGAGAGGACCAAGACTTTAAATTAATAATAATAATAATAATAATA 30430
QY 1006 ATCTTAGAAGTTATACCTTG-CATTTTCTCGGAAATGAGAACTACCTTTTATATGT 1064
DB 30429 ATCTTAGAATTTATACCTTGAAATTTTCTCGGAAATGAGAACTACCTTTTATATGT 30370
QY 1065 GTATG-TTTTTATGCAATTAGCATTTATCTCTGTTTCAGGGAATATCTTTCCCTAAAGCA 1123
DB 30369 ACATGATTTTATGCAATTAACATTTGATCTCTGTTGAAGAAATGATTTCTTAAGACA 30310
QY 1124 AATAGTTAGATATAAGATTAAATTAATAATAATAATAATAATAATAATAATAATAATA 1158
DB 30309 ATAATGTTGGATATGAAGATTAGAACTTAACATA 30275

RESULT 9
AC017005/C
LOCUS      Homo sapiens BAC clone RP11-100J21 from Y, complete sequence.
DEFINITION AC017005
ACCESSION  AC017005.6 GI:13435269
VERSION    AC017005.6
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 161879)
AUTHORS    Sulston, J.E. and Waterston, R.
TITLE      Toward a complete human genome sequence
JOURNAL    Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE    99063792
PUBMED     9847074
REFERENCE   2 (bases 1 to 161879)
AUTHORS    Armstrong, J., Stoneking, T., Hawkins, M. and Bernard, K.
TITLE      The sequence of Homo sapiens BAC clone RP11-100J21
JOURNAL    Unpublished
REFERENCE   3 (bases 1 to 161879)
AUTHORS    Waterston, R.H.
TITLE      Direct Submission
JOURNAL    Submitted (09-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4 (bases 1 to 161879)
AUTHORS    Waterston, R.H.
TITLE      Direct Submission
JOURNAL    Submitted (23-MAR-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA

```

## REFERENCE

5 (bases 1 to 161879)

## AUTHORS

Waterston,R.

## JOURNAL

Direct Submission

## COMMENT

Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Mar 23, 2001 this sequence version replaced gi:9211335.  
 ----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)

----- Summary Statistics

Center project name: H\_NH0100J21

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

## SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Ooeagawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-79J10; the clone sequenced to the right is RP11-245K4, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-100J21; actual end is at base position 66964 of RP11-245K4.

## FEATURES

source

1. .161879  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="y"  
 /map="y"  
 /clone\_lib="RPCT-11"

repeat\_region

1. .1185

/rpt\_family="L1"

repeat\_region

1183. .1513

/rpt\_family="L1"

repeat\_region

1679. .1842

/rpt\_family="MER1\_type"

repeat\_region

2125. .3135

/rpt\_family="ERV1"

repeat\_region

3116. .3412

/rpt\_family="ERV1"

repeat\_region

3450. .3866

/rpt\_family="ERV1"

repeat\_region

3867. .6977

repeat\_region  
 6986. .7295  
 /rpt\_family="ERV1"  
 repeat\_region  
 7297. .7625  
 /rpt\_family="ERV1"  
 repeat\_region  
 9507. .9570  
 /rpt\_family="MER1\_type"  
 11143. .11817  
 /rpt\_family="L1"  
 repeat\_region  
 12324. .12843  
 /rpt\_family="L1"  
 repeat\_region  
 12874. .12922  
 /rpt\_family="ERV1"  
 repeat\_region  
 12885. .12929  
 /rpt\_family="ERV1"  
 repeat\_region  
 13051. .13264  
 /rpt\_family="L1"  
 repeat\_region  
 13381. .13619  
 /rpt\_family="L1"  
 repeat\_region  
 13828. .14306  
 /rpt\_family="L1"  
 repeat\_region  
 15532. .15662  
 /rpt\_family="ERV1"  
 repeat\_region  
 15740. .16044  
 /rpt\_family="ERV1"  
 repeat\_region  
 16045. .16136  
 /rpt\_family="ERV1"  
 repeat\_region  
 16178. .16239  
 /rpt\_family="ERV1"  
 repeat\_region  
 16728. .17201  
 /rpt\_family="ERV1"  
 repeat\_region  
 17188. .17253  
 /rpt\_family="ERV1"  
 repeat\_region  
 17243. .17602  
 /rpt\_family="ERV1"  
 repeat\_region  
 17746. .18064  
 /rpt\_family="L1"  
 repeat\_region  
 18085. .18235  
 /rpt\_family="L1"  
 repeat\_region  
 18303. .18876  
 /rpt\_family="ERV1"  
 repeat\_region  
 18927. .19413  
 /rpt\_family="ERV1"  
 repeat\_region  
 19419. .19628  
 /rpt\_family="L1"  
 repeat\_region  
 21508. .21759  
 /rpt\_family="Alu"  
 repeat\_region  
 22725. .22868  
 /rpt\_family="ERV1"  
 repeat\_region  
 22868. .23335  
 /rpt\_family="ERV1"  
 repeat\_region  
 23350. .24076  
 /rpt\_family="ERV1"  
 repeat\_region  
 24273. .24690  
 /rpt\_family="ERV1"  
 repeat\_region  
 24702. .24823  
 /rpt\_family="ERV1"  
 repeat\_region  
 24886. .25175  
 /rpt\_family="Alu"  
 repeat\_region  
 25186. .27543  
 /rpt\_family="ERV1"  
 repeat\_region  
 27654. .27697  
 /rpt\_family="ERV1"  
 repeat\_region  
 28186. .28421  
 /rpt\_family="ERV1"  
 repeat\_region  
 28440. .28804  
 /rpt\_family="ERV1"  
 repeat\_region  
 28773. .29267  
 /rpt\_family="ERV1"  
 repeat\_region  
 29270. .31939  
 /rpt\_family="ERV1"  
 repeat\_region  
 31933. .32211  
 /rpt\_family="ERV1"



**TITLE**  
**JOURNAL**  
 Submitted (25-APR-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
**REFERENCE**  
**AUTHORS**  
 7 (bases 1 to 166436)  
 Waterston, R.  
**TITLE**  
**JOURNAL**  
 Submitted (30-SEP-2000) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
**COMMENT**  
 On Apr 23, 2000 this sequence version replaced gi:6524293.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0506M09  
 -----

**NOTICE:** This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.

#### MAPPING INFORMATION:

The position of this clone was established as part of a  
 collaboration between the Human Chromosome Y Mapping Project  
 (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen,  
 and David C. Page at the Whitehead Institute for Biomedical  
 Research, Cambridge MA) and the Washington University Genome  
 Sequencing Center, St. Louis MO.

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male  
 donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frenken, E.,  
 Tatenoe, M., Catanese, J. J., and de Jong, P. J. (1998). An improved  
 approach for construction of bacterial artificial chromosome  
 libraries. *Genomics* 51:1-8. The clone may be obtained either from  
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
 and coworkers at the Roswell Park Cancer Institute  
 (<http://bacpac.med.buffalo.edu>)  
 VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-221K4. Actual start of  
 this clone is at base position 1 of RP11-506M9; actual end is at  
 base position 166436 of RP11-506M9.

#### FEATURES

source  
 1..166436  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="Y"  
 /map="y"  
 /clone="RP11-506M9"  
 /clone\_lib="RPCI-11"  
 195..923  
 /rpt\_family="L1"  
 repeat\_region  
 924..1077  
 /rpt\_family="L1"  
 repeat\_region  
 1060..1830  
 /rpt\_family="L1"  
 repeat\_region  
 1879..1918  
 /rpt\_family="L1"  
 repeat\_region  
 1905..3257  
 /rpt\_family="L1"  
 repeat\_region  
 3260..3393  
 /rpt\_family="Alu"  
 repeat\_region  
 3618..3694  
 /rpt\_family="MIR"  
 repeat\_region  
 4250..4511  
 /rpt\_family="Alu"  
 repeat\_region  
 4512..5804  
 /rpt\_family="L1"  
 repeat\_region  
 5804..7682  
 /rpt\_family="L1"  
 repeat\_region  
 8188..8294  
 /rpt\_family="L2"  
 repeat\_region  
 9429..9499  
 /rpt\_family="L2"  
 repeat\_region  
 9826..9872  
 /rpt\_family="(A)n"  
 repeat\_region  
 9962..10035  
 /rpt\_family="L2"  
 repeat\_region  
 10596..10684  
 /rpt\_family="CA-rich"  
 repeat\_region  
 12836..13209  
 /rpt\_family="MaLR"  
 repeat\_region  
 14667..14689  
 /rpt\_family="AT-rich"  
 repeat\_region  
 16692..16712  
 /rpt\_family="AT-rich"  
 repeat\_region  
 17516..17584  
 /rpt\_family="AT-rich"  
 repeat\_region  
 17590..17667  
 /rpt\_family="MERS3"  
 repeat\_region  
 17814..18112  
 /rpt\_family="MERS3"  
 repeat\_region  
 19421..19650  
 /rpt\_family="Alu"  
 repeat\_region  
 19889..19813  
 /rpt\_family="L1"  
 repeat\_region  
 19815..20107  
 /rpt\_family="Retroviral"  
 repeat\_region  
 20410..20861  
 /rpt\_family="Retroviral"  
 repeat\_region  
 21083..25178  
 /rpt\_family="Retroviral"  
 repeat\_region  
 25196..27960  
 /rpt\_family="Retroviral"  
 repeat\_region  
 28006..28520  
 /rpt\_family="Retroviral"  
 repeat\_region  
 28328..28905  
 /rpt\_family="Retroviral"  
 repeat\_region  
 29157..29395  
 /rpt\_family="L2"  
 repeat\_region  
 29396..29438  
 /rpt\_family="Alu"  
 repeat\_region  
 30965..30996  
 /rpt\_family="(TAA)n"  
 repeat\_region  
 31776..32149  
 /rpt\_family="AT-rich"  
 repeat\_region  
 33205..33489  
 /rpt\_family="MaLR"  
 repeat\_region  
 35802..35832  
 /rpt\_family="Alu"  
 repeat\_region  
 36311..36389  
 /rpt\_family="AT-rich"  
 repeat\_region  
 36757..36827  
 /rpt\_family="(TTAAA)n"  
 repeat\_region  
 38147..38551  
 /rpt\_family="MERS3"  
 repeat\_region  
 38552..38576  
 /rpt\_family="Retroviral"  
 repeat\_region  
 38577..38633  
 /rpt\_family="(TC)n"  
 repeat\_region  
 38634..39399  
 /rpt\_family="Retroviral"  
 repeat\_region  
 39400..39459  
 /rpt\_family="Retroviral"

```
/rpt_family="MER1_type"
39460. .40129
/rpt_family="Retroviral"
40121. .41902
/rpt_family="Retroviral"
41931. .42280
/rpt_family="Retroviral"
42281. .42759
/rpt_family="Retroviral"
42961. .43305
/rpt_family="L1"
43410. .43703
/rpt_family="Alu"
44044. .44848
/rpt_family="L1"

Query Match      58.1%; Score 682.8; DB 9; Length 166436;
Best Local Similarity 81.4%; Pred. No. 6.4e-126;
Matches 904; Conservative 0; Mismatches 162; Indels 44; Gaps 8;

QY 59 AGGGTCTTGAAGCTTTGAGATTAACAATGGCAGGAAATCATCTTTTAAAGTAAT 118
Db 148862 AGGGTCTTGAAGCTTTGAGATTAACAATGGCAGGAAATCATCTTTTAAAGTAAT 148921

QY 119 TCTCCTTGGAGATGGTGGAGTGGGAGAGTTCACCTATGAACAGATATGTAACATAA 178
Db 148922 TCTCCTTGGAGATGGTGGAGTGGGAGAGTTCACCTATGAACAGATATGTAACATAA 148981

QY 179 GTTTGATACCCAGCTCTTCATACATAGGTGTGGAAATTTTAAATAAAGATTGCAAGT 238
Db 148982 GTTTGATACCCAGCTCTTCATACATAGGTGTGGAAATTTTAAATAAAGATTGCAAGT 149039

QY 239 GATGACATTTTGTACCATGACAGATTTGGGACACGGCAGGTGAGGACGATTCGGAAG 298
Db 149040 GATGACATTTTGTACCATGACAGATTTGGGACACGGCAGGTGAGGACGATTCGGAAG 149099

QY 299 CTTGAGACACCACTTTTACAGAGTCTGACTGCTGCTGCTTACTTTTGTAGTGTGATGA 358
Db 149100 CTTGAGACACCACTTTTACAGAGTCTGACTGCTGCTGCTTACTTTTGTAGTGTGATGA 149159

QY 359 TTCACAAAGCTCCAGAACTTAAGTAACCTGGAAGAAAGAAATTCATATATATGTCAGATGT 418
Db 149160 TTCACAAAGCTCCAGAACTTAAGTAACCTGGAAGAAAGAAATTCATATATATGTCAGATGT 149189

QY 419 GAAAGAGCTGAGAGCTTCCCTTTGTGATTTCTGGTACACAGATTCATAGGACGAGC 478
Db 149190 GAAAGAGCTGAGAGCTTCCCTTTGTGATTTCTGGTACACAGATTCATAGGACGAGC 149249

QY 479 GCAGTGTCTACAGAGAACCCCAAGCTTGTGTCAGGACCAACGGCGACTATCCTTTATTT 538
Db 149250 GCAGTGTCTACAGAGAACCCCAAGCTTGTGTCAGGACCAACGGCGACTATCCTTTATTT 149308

QY 539 TGAACAGTGCMAAGATGCCACAAATGTGGCAGCAGCTTTGAGGAGAGCGTTTCGAAG 598
Db 149309 TGAACAGTGCMAAGATGCCACAAATGTGGCAGCAGCTTTGAGGAGAGCGTTTCGAAG 149368

QY 599 AGTTCTTGTACGAGGATAGTGCAGATCATTGATTCAGACAGACACAGTCAATCTTCA 658
Db 149369 AGTTCTTGTACGAGGATAGTGCAGATCATTGATTCAGACAGACACAGTCAATCTTCA 149428

QY 659 CCGAAGCCCAAGCTTGTGATTTCTGCTGTTGATTTAGTGGTGGTGGTGGTGGTGGTGGTGG 714
Db 149429 CCGAAGCCCAAGCTTGTGATTTCTGCTGTTGATTTAGTGGTGGTGGTGGTGGTGGTGGTGG 149488

QY 715 TAACCAACTCACACATATACAAATCAACATGAGTGGGATGGGAGAGAAATAGGTTTG 774
Db 149489 TAACCAACTCACACATATACAAATCAACATGAGTGGGATGGGAGAGAAATAGGTTTG 149548

QY 775 CAGCAGTGTATCATCTTACTATAAATAATTAACATAATGT-----TGCTGCTTCTATTAGTTGG 830
Db 149549 TAGCAGTGTATCTTACTATAAATAATTAACATAATGT-----TGCTGCTTCTATTAGTTGG 149608

QY 831 TGGGAGAGGGACATCCACTCTCTGGAGGAATATATTTACTCAATATGGCACCTTACA 890
```

```
Db 149609 TGGGAGATGGACATATCCACTCATGGAGATATATTTCTCAATATGAACCTTACA 149668
QY 891 TTTATAAATTGTAAAGTGTCTAATAACGTTTCTTTTAATTAATAATGTAAGTTGCAGA 950
Db 149669 TTTATAAATTGTAAAGTGTCTAATAAAGTCTCTTTGAATTACACATGCAATTAGAGA 149727
QY 951 GCTAATAATGAATGACCAAGACTTTAATATATATTAATAAATAAAGAACTTGACTATTCT 1010
Db 149728 GCTAATAATGAGAGGACCAAGACTTTAATATATATTAATAAATAAAGAACTTGACTATTCT 149787
QY 1011 AGAAGTTATATCTTG-GATTTTCTCGGAAATGGAAGAACTACTTTTATATATGTAATG 1069
Db 149788 AGAATTTATATCTTGAAATTTTCTCGGAAATGGAAGAACTACTTTTATATGTAATG 149847
QY 1070 -TTTTTATGCAATTAGCAATGTAATCTTCTGTCAGGAAATACTTTCTTAAAGCAATAAT 1128
Db 149848 ATTTTGTGCAATTAACATGTAATCTTCTGTCAGGAAATGTAATTTCTTAAAGCAATAAT 149907
QY 1129 GTTAGATATTAAAGATTAAATCTAATGTA 1158
Db 149908 GTTAGATTAAGAACTTAGAATCTAATGTA 149937

RESULT 11
AC007965/c
LOCUS
DEFINITION Homo sapiens BAC clone RP11-245K4 from Y, complete sequence.
ACCESSION AC007965
VERSION AC007965.3 GI:10047995
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 182083)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 182083)
AUTHORS Du, H., Stoneking, T. and Gregory, S.
TITLE The sequence of Homo sapiens BAC clone RP11-245K4
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 182083)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 182083)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 182083)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 10, 2000 this sequence version replaced gi:5523821.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0245K04
-----
NOTICE: This sequence may not represent the entire insert of this
```

clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skalecky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Ooegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

#### VECTOR: pBAC3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-100J21; the clone sequenced to the right is RP11-270H4, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-245K4; actual end is at base position 15260 of RP11-270H4.

The sequence RP11-245K4 from base positions 105307 to 105967 and 141470 to 141790 is represented by sequence derived from PCR of clone dna.

#### FEATURES

```

source
    1..182083
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="Y"
        /map="y"
        /clone="RP11-245K4"
        /clone_lib="RPCI-11"
repeat_region
    1..65
        /rpt_family="L1"
repeat_region
    3688..3775
        /rpt_family="L1"
misc_feature
    9345..9559
        /note="similar to EST A1654951 (NID:G4738930) wb65g09.x1"
misc_feature
    9372..9555
        /note="similar to EST A1767654 (NID:G5234163) wh38b10.x1"
misc_feature
    10775..10876
        /note="similar to EST A1767654 (NID:G5234163) wh38b10.x1"
misc_feature
    11063..11277
        /note="similar to EST A1767654 (NID:G5234163) wh38b10.x1"
repeat_region
    11875..12075
        /rpt_family="MER1_type"
misc_feature
    12065..12386
        /note="similar to EST B5093625 (NID:G8484077)"
repeat_region
    12856..13310
        /rpt_family="MaLR"
repeat_region
    13650..13896
        /rpt_family="Alu"
repeat_region
    13919..14086
        /rpt_family="MER1_type"
repeat_region
    14854..14991
        /rpt_family="MER1_type"
repeat_region
    15334..15397

```

```

/rpt_family="L2"
15406..15530
/rpt_family="MIR"
15762..15841
/rpt_family="Alu"
15840..15972
/rpt_family="Alu"
16714..16830
/note="similar to EST AA020179 (NID:G1483961) mh51g02.r1"
16716..16830
/note="similar to EST AA874484 (NID:G2979173) vx80h07.r1"
16716..16829
/note="similar to EST A1918053 (NID:G5637908) w138g01.x1"
16716..16829
/note="similar to EST AW207673 (NID:G6507169)"
16716..16829
/note="similar to EST AW770235 (NID:G7702298) h179b07.x1"
16716..16829
/note="similar to EST T89537 (NID:G718050) ye04c07.r1"
16717..16830
/note="similar to EST A1928109 (NID:G5664073) wp11b12.x1"
16717..16830
/note="similar to EST AW140982 (NID:G6160750)"
16717..16829
/note="similar to EST AW027866 (NID:G5886622) ws62b12.x1"
16722..16830
/note="similar to EST AA799342 (NID:G4131386)"
16736..16830
/note="similar to EST A1819740 (NID:G5438819) wj11g03.x1"
16736..16829
/note="similar to EST A1312610 (NID:G4018215) qp77b06.x1"
16751..16829
/note="similar to EST N40075 (NID:G1163620) yx98e03.r1"
18145..18569
/rpt_family="MaLR"
18606..19251
/note="similar to EST AA203531 (NID:G1799242) zx56g09.r1"
18782..18932
/note="similar to EST AA799342 (NID:G4131386)"
18782..18932
/note="similar to EST AA854621 (NID:G2942159) aj76f06.s1"
18782..18932
/note="similar to EST AA874484 (NID:G2979173) vx80h07.r1"
18782..18932
/note="similar to EST A1819740 (NID:G5438819) wj11g03.x1"
18782..18932
/note="similar to EST A1928109 (NID:G5664073) wp11b12.x1"
18782..18932
/note="similar to EST AW140982 (NID:G6160750)"
18782..18929
/note="similar to EST AA020179 (NID:G1483961) mh51g02.r1"
18784..18932
/note="similar to EST AA715483 (NID:G2727757) nv53f10.r1"
18784..18932
/note="similar to EST A1312610 (NID:G4018215) qp77b06.x1"
18784..18932
/note="similar to EST A1918053 (NID:G5637908) w138g01.x1"
18784..18932
/note="similar to EST AW027866 (NID:G5886622) ws62b12.x1"
18784..18932
/note="similar to EST AW207673 (NID:G6507169)"
18784..18932
/note="similar to EST AW770235 (NID:G7702298) h179b07.x1"
18784..18932
/note="similar to EST N40075 (NID:G1163620) yx98e03.r1"
18784..18932
/note="similar to EST T89537 (NID:G718050) ye04c07.r1"
18798..18932
/note="similar to EST A1033300 (NID:G3254253) ow74e10.s1"
19378..19473
/note="similar to EST AA799342 (NID:G4131386)"
19378..19473
/note="similar to EST T89537 (NID:G718050) ye04c07.r1"

```



```
misc_feature 19378..19463
/notes="similar to EST AA715483 (NID:92727757) nv53f10.r1"
misc_feature 19378..19463
/notes="similar to EST AA854621 (NID:92942159) aj76f06.s1"
misc_feature 19378..19463
/notes="similar to EST AI312610 (NID:94018215) qp77b06.x1"
misc_feature 19378..19463

Query Match
Best Local Similarity 81.4%; Pred. No. 6.2e-126; Length 182083;
Matches 904; Conservative 0; Mismatches 162; Indels 44; Gaps 8;

QY 59 AGGTTCTTGAAGTTTGGAGTAAACAATGGCAGGAAATCATCACTTTTAAAGTAAT 118
Db 31142 AGGGTTCTTGAAGCTTTCGAATTAACAGTGGCAGGAAATTCATCACTTTTAAAGTAAT 31083
QY 119 TCTCCTTGGAGATGGTGGAGTTCACCTATGACAGATATGTAACTAATAA 178
Db 31082 TCTCTTGGAGATGGTGGAGTTCACCTATGACAGATATGTAACTAATAA 31023
QY 179 GTTGTATACCCAGCTCTTCATACATAGGTGTGGAATTTTAAATAAGATTTGGAAGT 238
Db 31022 GTTGTGCACTCAGCTGTTCATCAATAGATGTGGAATTTTAAATAA--GATTTGAAG 30965
QY 239 GGATGGACATTTTGTACCATGAGATTTGGGACAGGAGTTCAGGACGATTCGGAAG 298
Db 30964 GGTTGGACATTTTGTACCATGAGATTTGTGACATGGCAGGTCAGGAGATTTCTGAAG 30905
QY 299 CTTGAGGACACCATTTTACAGAGTTCTGACTGCTGCTTACTTTTGTAGTGTGATGA 358
Db 30904 CTTGAGGACACCATTTTACAGAGTTCTGACAGTTCTGCTTATTTTGTAGATCAATGA 30845
QY 359 TCCAAAGCTTCCAGACTTAAGTAACCTGGAAGAAAGAAATTCATATATATGCGAGATGT 418
Db 30844 TTTAGGAAGCTTCCAGAACTTAAGTAATG----- 30815
QY 419 GAAAGAGCTTGAGAGCTTTCCTTTTCTGATCTGGGTAAACAAGTTGACATGAAGCGAGC 478
Db 30814 GAAAGAGCTTGAGAGCTTTCCTTTTCTGATCTGGGTAAACAAGTTGACATGAAGCGAGC 30755
QY 479 GCAGGTGTCTACAGAAAGACCCCAAGCTTGGTGACGGGACACGGCGACTATCCTTTATTT 538
Db 30754 GCAGGTGTCTTATAGAGAA--CAGCTGCATTTGACAGAAAGGACAGTGACTACCTTACTT 30696
QY 539 TGAACAAGTGCAAAAGATGCCAATAATGTGGCAGCAGCTTTGAGGAAGCGGTTGCGAAG 598
Db 30695 TGAACAAGTGCAAAAGATGCCAATAATGTGGCAGCAGCTTTGAGGATGGGGTTCAAAA 30636
QY 599 AGTTCTTGCTACCGAGATAGGTTCAGATCATTTGATTCAGACAGACACAGTCAATCTTCA 658
Db 30635 AATCTTGTGACTGTGGATATGTCAGATCATTTGATGAACACAGGCGATGGTCAACCTTCA 30576
QY 659 CCGAAAGCCCAAGCCCTAGCTCATCTTCTGTTGATTTAG-----ATTGTTGATCATTC 714
Db 30575 CCAAAAGCCCAAGCCCTGGCAGATCTTGTATTGATTTAGATTAAGGATGAAGCAGATGCATGC 30516
QY 715 TACCAACTCACACATATACAAAATCAACATGGGGATGGGAGAGAGAAATTAGGTTTG 774
Db 30515 TAACCAATTCACACATATACAAAATCAACATGGGATGGGAGAGACAAATTAGTGTGTTG 30456
QY 775 CAGCAGTGTATCATCTACTAATAAATAAATAAATAATG-----TGCTGCTTCATTAGTTGG 830
Db 30455 TACAGTGTATTACTTACTAATAAAGATTCACAAATGTATAGCCCTGTGTCATTAGTTGG 30396
QY 831 TGGGAGAGGACACATCCACTCTTGGAGGATATATTTACTCAATAATAGGCACCTTACA 890
Db 30395 TGGGAGATGGGACATATCCACTCATGGAGGAATATATTTCTCAATAATGAACCTTACA 30336
QY 891 TTTATAAATTTGAACAGTTGTCTTAATAAGCTTTCTTTTAAATTTAAATATCTAAGTTGCGA 950
Db 30335 TTTATAAATTTGAATGGTTGTCTTA--AAAGTCTCTTTGATTTTACATGCAATTTAGAGA 30277
QY 951 GCTAATAAATGAATGACCAAGCACTTTAATTAATAAATAAATAAATAAAGAACTTGACTATTCT 1010
```

```
Db 30276 GCTAATAAATGAGAGGACCAAGACTTTAATTAATTAATTAAGATAAGACTTACTTCT 30217
QY 1011 AGAGTTATATCTTG-GATTTTCTCGGAAAATGGAGAACTACTTTTATATGTGTATG 1069
Db 30216 AGAATTTATATCTTGAATTTTCTCGGAAAATGGGAAAATAATTTTCTATGTACATG 30157
QY 1070 -TTTTATGCAATAGCAATGTATTTCTGGTTTCAGGAAAATACTTTCTTAAAGCAATAAT 1128
Db 30156 ATTTTGTGCAATTAACATTTGTATTTCTGGTGAAGAAATGATTTCTTAAGACAATAAT 30097
QY 1129 GTTAGATATTAAGCAATTAATAATCTAATGTA 1158
Db 30096 GTTGTATATGAAGATTAGAATCTTAACATA 30067

RESULT 12
BC008160
LOCUS
DEFINITION
Mus musculus RAB9, member RAS oncogene family, mRNA (CDNA clone
MGC:11411 IMAGE:3964134), complete cds.
ACCESSION
BC008160
VERSION
BC008160.1 GI:14919427
KEYWORDS
MGC.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1251)
Strausberg,R.L., Peingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Sheetz,T.E., Brownstein,M.J., Ustin,T.B., Teshituki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1251)
Direct Submission
Strausberg,R.
Submitted (22-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
TITLE
human and mouse cDNA sequences
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Naravati,
A.N., Gibbs, R.A.
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IRAC Place: 17 Row: K Column: 20  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 9790226.

# FEATURES

Location/Qualifiers  
 1..1251  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVE/N"  
 /db\_xref="taxon:10090"  
 /clone="MGC:11411 IMAGE:3964134"  
 /tissue type="mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."  
 /clone.lib="NCI CGAP\_Mam1"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV-SPORT6"  
 1..1251  
 /gene="Rab9"  
 /note="synonyms: Sid6061p, 2410064E05Rik"  
 /db\_xref="LocusID:56382"  
 /db\_xref="MGI:1890695"  
 186..791  
 /codon\_start=1  
 /product="RAB9, member RAS oncogene family"  
 /protein\_id="RAH08160.1"  
 /db\_xref="GI:14919428"  
 /db\_xref="LocusID:56382"  
 /translation="MAGKSSLPKILLDGVGVKSSLMNRVTVNPKDSLPHTIGVEF LNKLEVDVGHVFTWQINDTAGQERPSRLRFFYRGSBCLLTFSVDSDSQFNLNWK KEFIYADVPEFPFVILGNKTDIKERQVSTEEAQAQWCKNDGYPIFYTSKNDSTN VAAAFEEAVRILATEDSEHLIQTDTVLRHKPKNSOC"  
 390 a 247 c 270 g 344 t

## gene

## CDS

Query Match 56.6%; Score 664.6; DB 10; Length 1251;  
 Best Local Similarity 78.5%; Pred. No. 7.1e-122; Indels 30; Gaps 7;  
 Matches 893; Conservative 0; Mismatches 214;  
 QY 36 TGCATCTCTCTCAACACCCCTAGGAGGGTCTCTGAAGCTTTTCAGATTAAACAATGGCAGGA 95  
 DB 135 TGCCTGAAGTGAAGTCTGAAGGATTCGAATCAAACTTACATTATTAATGAAGTGAAG 194  
 QY 96 AAATCATCATCTTTTAAAGTAATCTCTCTGAGAGTGTGGAGTTGGAGAGGTTCACTT 155  
 DB 195 AAATCGTCTCTTTTAAATTAATCTCTCTGAGAGTGTGGAGTTGGAGAGGTTCTCTT 254  
 QY 156 ATGAACAGATATGTAAGTAAAGTTTGATACCCAGCTCTTCCATACAAATAGGTGTGAA 215  
 DB 255 ATGAACAGATATGTAAGTAAATTAATTAATGATTCCTCCAGCTCTTCCACAAATAGGTGTGAA 314  
 QY 216 TTTTAAATAAAGATTTGGAAGTGGATGACATTTTGTATCCATGACAGATTGGGACAG 275  
 DB 315 TTTTAAATAAAGATCTGGAGGTGGAGGACGACATTTTGTATCCATGACAGATTGGGACACA 374  
 QY 276 GCAGTCTAGGAGCGATTCGAGCGCTGAGGACACCATTTTACAGAGGTTCTGACTGTGC 335  
 DB 375 GTGGCCAGAGACCGTCTCGAAGCGCTGAGGAGCGCATTTTACCAGGTTCTGACTGTTC 434  
 QY 336 CTGCTTACTTTTGTGTCGATGATTCACAAAGCTTCCAGAACTTAAAGTAAGTGAAGAAA 395  
 DB 435 CTGCTTACTTTTGTGTCGATGATTCACAGAGCTTCCAGAACTTGAAGCACTGGAAGAAA 494  
 QY 396 GAATTCATATATATGACAGATGTAAGAGCTGAGAGCTTCTCTTTTGTGATCTGGGT 455  
 DB 495 GAGTTTCATATATGACAGATGTAAGAGCGCTGAAAGCTTTCCTCTTTGTGATTTGGC 554  
 QY 456 AACAGATTGACATAGCAAGCGAGGTTGCTACAGAGAGAGCCCAAGCTTGGTGGCAGG 515  
 DB 555 AACAGACTGACATAGCAAGCGAGGTTGCTACAGAGAGAGCCCAAGCTTGGTGGCAGG 614  
 QY 516 GACAAAGCGGCGATCTCTTATTTTGAACAAGTGAAGAGATGCCACAAATGTGGCAGCA 575

DB 615 GACAAAGCGGCGATCTCTTACCTTTGAAACAAGTGCAGAAAGATTCACCAATGTCTGCT 674  
 QY 576 GCCTTTGAGGAAGCGGTTCGAAGAGTTCTTGTACCGAGGATAGTCAGATCATTTGATT 635  
 DB 675 GCCTTTGAGGAGCGAGTTCGAAGAATTTCTGGCCAGGAAGATAGTCAGAACACCTGATT 734  
 QY 636 CAGACAGACACAGTCAATCTTCACCGAAGCCCAAGCCCTAGCTCATCTTGTCTGTTGATTG 695  
 DB 735 CAGACAGACACAGTCAATCTTCACCGAAGCCCAAGCCCAAGCTCATCTTGTCTGTTGATTG 794  
 QY 696 -----TTAGATTGTTGATGCAATCTTAACCAACTCACATATACACAAAATCAACATGG 749  
 DB 795 CATTAAGGAGATAGTTGGTGCATCTTAACCAACACACACACACACAC-----ACAGAA 847  
 QY 750 GGATGGAGAGAGAAATTAGCGTTTCAGCAGGTGATCATCTACTACTAATA-AAATTAACCTA 808  
 DB 848 AGGTGGAGAAACAATTAGTATGCGAGAGAGTTCAATTTACTAATAAATTCAGTTAAGCG 907  
 QY 809 ATGTTGCTCTCTTCATTAGTTGGTGGAGAGGAGACATCCACTCTTGGAGGAATATATT 868  
 DB 908 ATATTGTTGCTCTCAATTAGTCGGTGGAGAGGACACTCACTC---TGGAGGAATCTATT 964  
 QY 869 TACTCAATATGCGACCTTACATTTATTAATTTGTAACAGTTGTCTAATAACGTTTCTTTA 928  
 DB 965 TACTCAGTAATGCGACCTTCCACTTATAAATTTGTAATTTGTTGCTTAAT---GTTTCTTTA 1021  
 QY 929 ATTTAAATATGTAAGTTGCGAGAGCTTAATAAATGAATGACCAAGACTTTAATAATAATAA 988  
 DB 1022 AATTAACAATTT-----AGTGTCTAATAAGATGACCAAGACTGACATTACTGTAAATTC 1074  
 QY 989 AATAAGAAACTTGACTATCTAGAGT---TATCTGAGATTTTCTTCTGGGAAAATGG 1045  
 DB 1075 AGACGACAACTTGGCTATTTAGAGTACACTTAGATTGTTTGGACCCACAGGAAATG 1134  
 QY 1046 AGAAGTACTTTTATATGTTATGTTTATCAATAGCTATGTTCTTCTGTTTCAGGG 1105  
 DB 1135 GAAATATCTTTTATATGTTATATTTTATGTAATTTATGTAATTTAGCATTTGTTCTAGGA 1194  
 QY 1106 AAATACTTTCTTAAAGCAATAATGTTAGATATTAAAGATTAATAATCTAATGTAAAAA 1162  
 DB 1195 AAGTAAATTTCTACAGCAGTACTATTAAAGATTAATAATCTAATAAATAAATAAATAA 1251

## RESULT 13

AB027290  
 LOCUS  
 DEFINITION  
 Mus musculus SID 99 mRNA for small GTP binding protein, complete cds.  
 1276 bp mRNA linear ROD 29-SEP-1999

## ACCESSION

AB027290

VERSION

AB027290.1 GI:5931615

KEYWORDS

small GTP binding protein.

SOURCE

Mus musculus (house mouse)

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 1276)

AUTHORS

Seki,N., Hattori,A., Hayashi,A., Kozuma,S., Muramatsu,M. and Saito,T.

TITLE

Mouse small GTP binding protein

JOURNAL

Published Only in DataBase (1999)

REFERENCE

2 (bases 1 to 1276)

AUTHORS

Seki,N., Hattori,A., Hayashi,A., Kozuma,S., Muramatsu,M. and Saito,T.

TITLE

Direct Submission

JOURNAL

Submitted (14-MAY-1999) Toshiyuki Saito, National Institute of Radiological Sciences, Genome Research Group; Inage-ku Anagawa 4-9-1, Chiba, Chiba 263-8555, Japan [E-mail:t\_saito@nirs.go.jp, Tel:81-43-201-3135, Fax:81-43-251-9818]

FEATURES

Location/Qualifiers

1..1276

/organism="Mus musculus"

/mol\_type="mRNA"

```

/db_xref="taxon:10090"
1. .1276
/gene="SID 99"
216. .821
/gene="SID 99"
/codon_start=1
/product="small GTP binding protein"
/protein_id="BAA84709.1"
/db_xref="GI:5931616"
/translation="MACKSLFKILLGDSGVKSLMRYVTKEDSOLFHTIGVEF
LNKDLVDGHPVMOIWDTAGQERFSLRFPYRGDCCLLTFSDVDSQFQLSNWK
KEFLIYADVPEPSFPVILGNKIDIKEROVSTEEBAQWCKONGDYPYFETSAKSTN
VAAFEAVARILLRSESHLIQTDVNLHRKPKNSCC"
BASE COUNT      384 a      264 c      274 g      354 t
ORIGIN

Query Match      56.2%; Score 660.2; DB 10; Length 1276;
Best Local Similarity 78.9%; Pred. No. 5.3e-121;
Matches 899; Conservative 0; Mismatches 203; Indels 37; Gaps 8;

QY 36 TGCATCTTCTCAACACCCCTAGGAGGTTCTTGAAGCGTTTGGAGATTAAACAATGGCAGGA 95
DB 165 TGCCTTGAAGTGAACCTACTAGGATTCGAATCAAACTTACATTATTAAATAGGCAGGA 224
QY 96 AATCATCATCTTTTAAAGTAATCTCTTGGAGATGGTGGAGTGGGAAGAGTTCACCT 155
DB 225 AATCGTCTCTTTTAAATAATCTCTTGGAGATGGTGGAGTGGGAAGAGTTCCTT 284
QY 156 ATGAACAGATGATGAACTAATAGTTTGATACCCAGCTCTTCCATACAAATAGTGTGGAA 215
DB 285 ATGAACAGATGATGAACTAATAGTTTGATACCCAGCTCTTCCACACATAGTGTGGAA 344
QY 216 TTTTAAATAAGATTGGAGTGGATGGACATTTTGTACCAATGAGATTTGGGACAG 275
DB 345 TTTTAAATAAGATCTGGAGTGGAGCGACATTTTGTACCAATGAGATTTGGGACAG 404
QY 276 GCAGGTGAGGAGCATTCCGAAGCTCGAGACACCATTTTACAGAGTCTTCTGACTGCTGC 335
DB 405 GCTGCGCCAGAACGCTTCCAGCTCGAGAGCGCATTTTACCGGGTCTTCTGACTGCTGC 464
QY 336 CTGCTTACTTTTGTGTCGATGATCAAAAGCTTCAGAACTTAAAGTAACTGGAAGAAA 395
DB 465 CTGCTTACTTTTGTGTCGATGATTCACAGAGCTTCAGAACTTAAAGTAACTGGAAGAAA 524
QY 396 GAATTCATATATTATGATGATGAAAGAGCTGAGAGCTTCTTCTTGTGATCTGGGT 455
DB 525 GAGTTTCAATATTACGAGATGAAAGAGCGCTTCCCTTTGTGATTTTGGGC 584
QY 456 AACAAAGATTGACATAAGCGAGCGAGTGTCTACAGAAAGCCCAAGCTTGGTGCAGG 515
DB 585 AACAAAGATTGACATAAGCGAGCGAGTGTCTACAGAAAGCCCAAGCTTGGTGCAGG 644
QY 516 GACACGGGCACTATCTTTATTTTGAACAAAGTGGCAAGATGCCCAAAATGTCGACGA 575
DB 645 GACACGGGCACTATCTTTATTTGAACAAAGTGGCAAGATGCCCAAAATGTCGCTCT 704
QY 576 GCCTTTGAGGAAGCGGTTCAAGAGTCTTGTACCGAGATAGGTCAGATCATTTGATT 635
DB 705 GCCTTTGAGGAGCGAGTTCAGAGATTTCTGCCAGGAGATAGGTCAGAACCTGATT 764
QY 636 CAGACGACACAGTCAATCTTCCAGGAAGCCCAAGCCCTAGCTTCCTTGTGTTGATTG 695
DB 765 CAGACGACACAGTCAATCTTCCAGGAAGCCCAAGCCCAAGCTTCCTTGTGTTGATTG 824
QY 696 -----TTAGATTGTTGATCTTCAACCACTGACATATACACAAATCAACATGG 749
DB 825 CATTAAGGGATAGTGTGATCTTCAACCAACACACACACACACACACACACACACAC 877
QY 750 GGATGAGAGAGAAATAGCGTTTGCAGAGTGTATCATCTACTATAA-AAATTAACCTA 808
DB 878 AGGTGAGAGAAACAAATAGTATGACAGAGTTCATTTTACTAATAAAATTCAGTTAACG 937
QY 809 ATGTTCTGCTTCAATAGTTGTTGGGAGAGGACACATCCACTCTTGTAGGAGATATATT 868
```

```

DB 938 ATATTGTTGCCTCATAGTCGTTGGAGAGGACACCTCCTC---TGAGGAATCTATT 994
QY 869 TACTCAATAATGGCACCTTACATTTTATAAATTTGAACAGTTGTCTTAATAACGTTTCTTTA 928
DB 995 TACTCAGTAATGGCACGTTCCACTTATAAATTTGAATTTGTTGCTAAAT---GTTTCTTTA 1051
QY 929 ATTAAATATGTAAGTTGCGAGAGCTAATAAATGAATGACCAAGACTTAAATATAATAAA 988
DB 1052 AATTAAACATTT-----AGTGCTAATAAGATGCCAAGAACTGACTTTTACTGTGAATTC 1104
QY 989 AATAAGAACTTGAATCTTCTAGAGT---TATACCTTGGATTTTCTCTGGGAAATGG 1045
DB 1105 AGACGACACCTTGGCTATTCTAGAGTACACTTAGATTGTTTGACCCACCAAGAAATG 1164
QY 1046 AGAACTACTTTTATATATGTTATGTTTATGCAATAGCATTTGTTCTTGTTCAGGG 1105
DB 1165 GAAATATTACTTTTATATGTTATATTTTATGTAATTTAGCATTTGTTCTTGGTCTAGGA 1224
QY 1106 AAATACTTTCTTAAAGCAATATGTTAGATATTAAGATTAATAATCTAATGTAAATAAA 1164
DB 1225 AAGTAAATCTTACAGCAGTA-----CTAATAAGATTAATAATCTAATGCAAAAAA 1276

AL672174      201642 bp      DNA      linear      ROD 09-NOV-2002
Mouse DNA sequence from clone RP23-408K4 on chromosome X, complete
sequence.
AL672174
ACCESSION     AL672174.8      GI:24940538
VERSION
KEYWORDS      HTG.
SOURCE        Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 201642)
Howden,P.
Direct Submission
Submitted (09-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 12, 2002 this sequence version replaced gi:21727414.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-408K4 is
from the RP23-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
Sequence from the Mouse Genome Sequencing Consortium whole genome
```

```
/organelle="Mus musculus"
/seq type="genomic DNA"
```

## NOTES

```
/organelle="Mus musculus"
/seq type="genomic DNA"
```

		/db_xref="taxon:10090"									
		/chromosomes="X"									
		/clone="RP23-316B4"									
		/clone_lib="RPC1-23"									
BASE COUNT		54830 a 37456 c 36136 g 56052 t									
ORIGIN											
Query Match											
Best Local Similarity 55.0%; Score 646.6; DB 10; Length 184474;											
Matches 901; Conservative 0; Mismatches 209; Indels 39; Gaps 9;											
QY	36	TGCATCTTCTCAACACCCCTAGGAGGTTCTTGAAGCTTTTGGAGATTAAACAATGGCAGGA	95								
Db	177038	TGCTTGAACCTGAACACTAATGATCTTAATCAAACTTACCTTATTATTAATGGCAGGA	177097								
QY	96	AAATCATCACTTTTAAAGTAATTCCTCTGGAGATGGTGGAGTTGGGAAGATTCACCTT	155								
Db	177098	AAATCGTCTCTTTTAAATAATTCCTCTGGAGATGGTGGAGTTGGCAAGATTCCTCTT	177157								
QY	156	ATGAACAGATATGTAACTAATAAGTTTGATACCCAGCTCTTCCATACAAATAGGTGGAA	215								
Db	177158	ATGAACAGATACGTAAACCAATTAATTTGATCCAGCTCTTCCACAAATAGGTGGAA	177217								
QY	216	TTTTTAATAAAGATTGGAAATGGATGGACATTTTGTATCCATGCAGATTTGGGACAG	275								
Db	177218	TTTTTAATAAAGATCTGGAGGTGGATGGACATTTTGTATCCATGCAGATTTGGGACACA	177277								
QY	276	GCAGGTCCAGAGCGATTCCGAAGCCTGAGGACACCATTTTACAGAGGTTCTGACTGCTGC	335								
Db	177278	GCTGGTCAAGAAGCATCTTGAAGCCTGAGGACACCATTTTACCGAGGTTCTGACTGTTGC	177337								
QY	336	CTGCTTACTTTTGTGTGATGATTAATCACAAAGCTTCCAGAACTTAAAGTAACCTGGAGAAA	395								
Db	177338	CTGCTTACTTTTGTGTGATGATTAATCACAAAGCTTCCAGAACTTAAAGTAACCTGGAGAAA	177397								
QY	396	GAATTCATATATATGACATGTTGAAGCCTGAGAGCTTCTTTTGTGATTTCTGGGT	455								
Db	177398	GAGTTTCATATATATGACATGTTGAAGCCTGAGAGCTTCTTTTGTGATTTTGGAC	177457								
QY	456	AACAGATTGACATAAGCCGAGCGGTGTTCTACAGAAAGCCCAAGCTTGGTGCAGG	515								
Db	177458	AACAGATTGACATAAAGAACCGGAGGTGTTCTACAGAAAGCCCAAGCTTGGTGCAGG	177517								
QY	516	GACACGGCGACTATCTTTATTTTGAACAAGATGCGAAAGATGCGCAAAATGGGCGACA	575								
Db	177518	GACAAATGGCGACTATCTTTATTTGAACAAGATGCGAAAGATGCGCAAAATGGTCTACT	177577								
QY	576	GCCTTTGAGGAAGCGTTCCGAAGATTTCTTGCTACGAGGATAGTTCAGATCATTTGATT	635								
Db	177578	GCCTTTGAGGAAGCGTTCTAGGAATTTGCGCCATGTAATATAGTTCAGAACACCTGATT	177637								
QY	636	CAGACAGACACAGTCAATCTTCAACGAAAGCCCAAGCCTAGCTCATCTTGTCTGTGATTG	695								
Db	177638	CAGACAGACACAGTCAATCTGACGAAAGCCCAAGCCTAGCTCATCTTGTCTGTGATTG	177697								
QY	696	-----TTAGATGTTGATGATTTCAACCACTCAGACATATACAAATCAACATGG	749								
Db	177698	CATTAAAGGATAGTTGGTCATTCTAACCAACACACACAGAGAGAA-----	177745								
QY	750	GGATGGAGAGAGAAATTAGCGTTTTCAGCAGTGTATCATCTACTAATA-AAAATAAAGTA	808								
Db	177746	-GGTGGAGAAACAATATGATGTCAGAGAGTTTCATCTACTAATAAATTCAGTTAACGC	177804								
QY	809	ATGTTGCTGCTTCATTTAGTTGGGAGAGGAGACATCCACTCTTGGAGGAATATATT	869								
Db	177805	ATATTGTTGCCCTCATTTAGTCGGTGGGAGGAGACATCACTC---TGGAGGAATCTATT	177861								
QY	869	TACTCAATAATGGCAGCTTACATTTTATAAATTTGAAGTGTCTTAATAACGTTCTTTTA	928								
Db	177862	TACTCAATATGGCAGCTTCCACTTATAAATTTGAATTTGTTGCTTAAT---GCTTCTTTA	177918								
QY	929	ATTAAATATGTAAGTTGAGAGCTTAATAATGAATGAAATGCAAGACTTTAATAATAA	988								

Search completed: February 10, 2004, 03:56:06  
Job time : 4605 secs

Db	177919	AAATTAAACATT-----AGTGCTAATAAGATGACCAAGCACTGACTTAAGTGTATTC	177971
QY	989	AAATAAGAAACCTTGACTATCTAGAAAGTTATACCTTGGATTTTTT-----CCTGGGAAAATG	1044
Db	177972	AGACACACACCTTGACTATCTAGAAAGTACACTTAGATTGTTTTTGACCCACAAAGGAATG	178031
QY	1045	GAGAACTTACTTTTTTATATGTATGTTTATGCAATTAGCATTGTATTTCTTGGTTTCAGG	1104
Db	178032	GAGAAATTAC-TTTTTATATGTATATATTTTTATGTAATTAGCATTGTATTTCTTGGTCTAGG	178090
QY	1105	GAATACTTTTCCCTAAAGCAATAATGTAGATATTAAGATTTAAATCTAATGTAAAAAAA	1164
Db	178091	AAAGTAAATTCCTACAGC-AGATAAAATCTAATAAGATTAAATTTAATGCAAAAAA	178149
QY	1165	AAAAAATAA 1173	
Db	178150	AAAAAATAA 178158	

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 23:51:41 ; Search time 390 Seconds  
(without alignments)  
8132.922 Million cell updates/sec

Title: US-09-988-974-6

Perfect score: 1175

Sequence: 1 CACTGTGATGAACACTTTT.....GTAAAAA.....1175

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1175	100.0	1175	20 AAX04755	CDNA encoding a Rab
2	1175	100.0	1175	24 ABQ75524	Human Rab protein
3	1160.2	98.7	24 ABQ75377	Human gene express	
4	1100	93.6	1969	22 ABA45112	Human breast cell
5	1100	93.6	1969	22 AAK03817	Human brain expres
6	1100	93.6	1969	22 AAI13885	Probe #3818 for ge
7	1100	93.6	1969	22 AAI03747	Probe #3738 used t
8	1100	93.6	1969	24 ABS03847	Human genome-deriv

9	1100	93.6	6058	22 AAL02805	Human reproductive
10	632	53.8	632	22 ABA50249	Human breast cell
11	632	53.8	632	22 AAK16569	Human brain expres
12	632	53.8	632	22 AAI23103	Probe #13036 for g
13	632	53.8	632	22 AAI08751	Probe #8742 used t
14	632	53.8	632	24 ABS16384	Human genome-deriv
15	604.4	51.4	606	25 ACA56484	Human signalling p
16	439.6	37.4	487	19 AAV61344	Extended CDNA seq
17	439.6	37.4	487	19 AAV58676	Prostate tumour sp
18	439.6	37.4	487	22 AAS63847	Human prostate cDN
19	439.6	37.4	487	22 AAS10198	Human prostate tum
20	439.6	37.4	487	22 AAH93555	Human prostate-spe
21	439.6	37.4	487	22 AAH84869	Human prostate-spe
22	439.6	37.4	487	22 AAH02620	Prostate tumour an
23	439.6	37.4	487	24 ABS71344	Prostate tumour tum
24	439.6	37.4	487	24 ABS58728	Prostate tumour CD
25	439.6	37.4	487	24 ABU95019	Human ID-4280 cDNA
26	439.6	37.4	487	25 ACA59456	Prostate cancer th
27	432.6	36.8	480	21 AAA06439	Human immunogenic
28	405.8	34.5	566	23 ABV59019	Human prostate exp
29	385.4	32.8	399	21 AAC30465	Human secreted pro
30	356	30.3	393	24 ABZ35026	Human gene express
31	333.2	28.4	1099	21 AAH96994	Nucleoside sequenc
32	333.2	28.4	1114	22 AAS27473	CDNA encoding nove
33	333.2	28.4	1135	24 ABQ60937	RAB9-like protein
34	328.4	27.9	1050	23 AAS91757	CDNA encoding nove
35	322.4	27.4	1111	22 AAS27054	CDNA encoding nove
36	306.2	26.1	473	19 AAV61314	CDNA sequence of p
37	306.2	26.1	473	19 AAV58583	Prostate tumour sp
38	306.2	26.1	473	21 AAA06346	Human immunogenic
39	306.2	26.1	473	22 AAS63554	Human prostate cDN
40	306.2	26.1	473	22 AAS10105	Human prostate tum
41	306.2	26.1	473	22 AAH93462	Human prostate-spe
42	306.2	26.1	473	22 AAH84776	Human prostate-spe
43	306.2	26.1	473	22 AAH02527	Prostate tumour an
44	306.2	26.1	473	24 ABS71351	Human prostate tum
45	306.2	26.1	473	24 ABS58635	Prostate tumour CD

## ALIGNMENTS

RESULT 1  
AAX04755  
ID AAX04755 standard; cDNA; 1175 BP.  
XX AAX04755;  
AC AAX04755 (first entry)  
DT 08-APR-1999  
XX cDNA encoding a Rab protein designated HRAB3-3.  
DE Rab protein; HRAB3-3; Incyte clone 641412; vesicle targeting;  
KW membrane transfer; protein processing; targeting; secretion;  
KW abnormal neurotransmitter production; lysosomal storage disease;  
KW choroideraemia; endocrine hypofunction; hyperfunction; ss.  
XX Homo sapiens.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 87..692  
FT /\*tag= a  
FT /product= HRAB33  
XX US5869291-A.  
XX 09-FEB-1999.  
XX 27-DEC-1996; 96US-0773423.  
XX 27-DEC-1996; 96US-0773423.  
XX (INCY-) INCYTE PHARM INC.

XX Bandman O, Goli SK, Hillman JL;  
 XX WPI; 1999-152769/13.  
 DR P-PSDB; AAW88105.  
 XX  
 XX DNA encoding human Rab proteins - useful for producing recombinant  
 PT proteins  
 XX  
 XX Example 1; Fig 3A-B; 36pp; English.  
 XX  
 CC The present sequence encodes Rab protein designated HRAB-3. The  
 CC nucleic acid sequence was first identified in Incyte clone 641412  
 CC from a breast tissue cDNA library. The HRAB nucleic acid sequences  
 CC are useful for producing recombinant human Rab proteins which may  
 CC be useful for treating disorders involving vesicle targeting,  
 CC membrane transfer or fusion, or protein processing, targeting or  
 CC secretion, e.g. disorders characterised by abnormal neurotransmitter  
 CC production and release, lysosomal storage diseases, chorioideraemia and  
 CC endocrine hypofunction and hyperfunction.  
 XX  
 XX Sequence 1175 BP; 387 A; 191 C; 244 G; 353 T; 0 other;  
 SQ  
 Query Match 100.0%; Score 1175; DB 20; Length 1175;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-244;  
 Matches 1175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CACTGTGATGAACACACTTTCCCGTGTCTGTTGAGTGCATCTTCTCAACAACCCCTAGGAG 60  
 DB 1 CACTGTGATGAACACACTTTCCCGTGTCTGTTGAGTGCATCTTCTCAACAACCCCTAGGAG 60  
 QY 61 GGTTCCTGAAGCTTTTGAAGATTAACAATGCGCAGGAAATCATCTTTTAAAGTAATTC 120  
 DB 61 GGTTCCTGAAGCTTTTGAAGATTAACAATGCGCAGGAAATCATCTTTTAAAGTAATTC 120  
 QY 121 TCCTTGGAGATGCTGAGTTGGAGAGTTCACTATGAACAGATATGTAATTAAGT 180  
 DB 121 TCCTTGGAGATGCTGAGTTGGAGAGTTCACTATGAACAGATATGTAATTAAGT 180  
 QY 181 TTGATACCCAGCTTCCATACATAGTGTGGAAATTTTAAATAAAGATTGGAAGTGG 240  
 DB 181 TTGATACCCAGCTTCCATACATAGTGTGGAAATTTTAAATAAAGATTGGAAGTGG 240  
 QY 241 ATGGACATTTTCTACCATGAGATTGGGACACGCGAGGTGAGGAGGATTCGAGCC 300  
 DB 241 ATGGACATTTTCTACCATGAGATTGGGACACGCGAGGTGAGGAGGATTCGAGCC 300  
 QY 301 TGAGGACACCATTTTACAGAGGTTCTGACTGCTGCTGCTTACTTTTAGTGTGCGATGATT 360  
 DB 301 TGAGGACACCATTTTACAGAGGTTCTGACTGCTGCTGCTTACTTTTAGTGTGCGATGATT 360  
 QY 361 CACAAAGCTTCCAGAACTTAAGTAATCTGGAAGAAAGATTTCATATATTATGCAATGTGA 420  
 DB 361 CACAAAGCTTCCAGAACTTAAGTAATCTGGAAGAAAGATTTCATATATTATGCAATGTGA 420  
 QY 421 AAGAGCTTCAGAGCTTTCTCTTTTGTGATTCTGGGTAAACAAGATTACATAAGCGAAGCGC 480  
 DB 421 AAGAGCTTCAGAGCTTTCTCTTTTGTGATTCTGGGTAAACAAGATTACATAAGCGAAGCGC 480  
 QY 481 AGGTGTCTACAGAAAGCCCAAGCTTGGTGGAGCAACGCGACTATCTTATTATTG 540  
 DB 481 AGGTGTCTACAGAAAGCCCAAGCTTGGTGGAGCAACGCGACTATCTTATTATTG 540  
 QY 541 AAACAAGTCAAAAGATGTCACAAATGTGGCAGAGCCCTTTGAGGAAGCGGTTCGAAGAG 600  
 DB 541 AAACAAGTCAAAAGATGTCACAAATGTGGCAGAGCCCTTTGAGGAAGCGGTTCGAAGAG 600  
 QY 601 TTCTTGCTACCGAGGATAGGTGAGATCAATTTGATTGAGACAGACACAGTCAATCTTACC 660  
 DB 601 TTCTTGCTACCGAGGATAGGTGAGATCAATTTGATTGAGACAGACACAGTCAATCTTACC 660  
 QY 661 GAAAGCCCAAGCTTAGCTCATCTTCTGTTGATTGTTAGATTGTTGATGCTTCAACCA 720  
 DB 661 GAAAGCCCAAGCTTAGCTCATCTTCTGTTGATTGTTAGATTGTTGATGCTTCAACCA 720

DB 661 GAAAGCCCAAGCTTAGCTCATCTTCTGTTGATTGTTAGATTGTTGATGCTTCAACCA 720  
 QY 721 ACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTGCAGCAG 780  
 DB 721 ACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTGCAGCAG 780  
 QY 781 TGTATCATCTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 840  
 DB 781 TGTATCATCTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 840  
 QY 841 GACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTATAAAT 900  
 DB 841 GACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTATAAAT 900  
 QY 901 GTAACAGTGTCTTAATAAAGTCTTTCTTTAATTTAAATATGTAAGTTCAGAGCTTAATAAT 960  
 DB 901 GTAACAGTGTCTTAATAAAGTCTTTCTTTAATTTAAATATGTAAGTTCAGAGCTTAATAAT 960  
 QY 961 GAAATGACCAAGACTTTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1020  
 DB 961 GAAATGACCAAGACTTTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1020  
 QY 1021 CTTGGATTTTCTCGGAAATGGAGAACTACTTTTATATGTTATGTTATGTTATGCTTATGCTAA 1080  
 DB 1021 CTTGGATTTTCTCGGAAATGGAGAACTACTTTTATATGTTATGTTATGTTATGCTTATGCTAA 1080  
 QY 1081 TTAGCATTTGTTCTTGGTTCAGGAAATACATTTCTTAAAGCAATAATGTTAGATATTAA 1140  
 DB 1081 TTAGCATTTGTTCTTGGTTCAGGAAATACATTTCTTAAAGCAATAATGTTAGATATTAA 1140  
 QY 1141 AGATTAAATCTAATGTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1175  
 DB 1141 AGATTAAATCTAATGTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1175  
 RESULT 2  
 ABQ75524  
 ID ABQ75524 standard; cDNA; 1175 BP.  
 XX  
 AC ABQ75524;  
 XX  
 DT 08-NOV-2002 (first entry)  
 XX  
 DE Human Rab protein HRAB-3 encoding cDNA SEQ ID NO:6.  
 XX  
 KW Human; Rab protein; HRAB-3; antianaemic; antidiabetic; antiinflammatory;  
 KW immunosuppressive; dermatological; neuroprotective; antiulcer; anti-HIV;  
 KW antirheumatoid; antiarthritic; cytostatic; endocrine; gene therapy;  
 KW immune system disorder; autoimmune haemolytic anaemia; myasthenia gravis;  
 KW juvenile diabetes mellitus; lupus erythematosus; multiple sclerosis;  
 KW Crohn's disease; rheumatoid arthritis; AIDS; ulcerative colitis; cancer;  
 KW leukaemia; lymphoma; carcinoma; vesicle targeting disease; ras family;  
 KW lysosomal storage disease; chorioideraemia; endocrine hypofunction;  
 KW endocrine hypofunction; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 1..2  
 FT FT /\*tag= a  
 FT FT /note= "not shown in Figure 1"  
 FT FT 87..692  
 FT FT /\*tag= b  
 FT FT /product= "HRAB-3"  
 FT FT /note= "Rab protein"  
 XX  
 FN US2002090712-A1.  
 XX  
 PD 11-JUL-2002.  
 XX  
 PF 19-NOV-2001; 2001US-0988974.  
 XX  
 PR 27-DEC-1996; 96US-0773423.

PR 17-DEC-1998; 98US-0215887.  
 XX (INCY-) INCYTE PHARM INC.  
 PA Hillman JL, Goli SK, Bandman O;  
 PI WPI: 2002-655836/70.  
 DR P-PSDB; ABP53078.  
 XX Novel Rab polypeptides, low molecular weight guanidine triphosphatases  
 PT of Ras superfamily, useful for treating, preventing or diagnosing  
 PT cancer, and immune system disorders e.g. ulcer, arthritis and Crohn's  
 PT disease -  
 XX Claim 5; Fig 3A-B; 40pp; English.  
 XX The present sequence encodes human HRAB-3 which is a Rab protein (low  
 CC molecular weight guanidine triphosphatases of the Ras superfamily).  
 CC Rab proteins (i) have anti-neuritic, antidiabetic, immunosuppressive,  
 CC anti-inflammatory, dermatological, neuroprotective, anti-rheumatoid,  
 CC anti-ulcer, anti-arthritis, anti-HIV, cytostatic and endocrine activities,  
 CC and can be used in gene therapy. (i) can be used for screening a compound  
 CC for effectiveness as an agonist or antagonist, for screening a compound  
 CC that specifically binds to (i) or modulates the activity of (i), and for  
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
 CC (i) can be used for diagnosing, treating or preventing immune system  
 CC disorders (including autoimmune haemolytic anaemia, juvenile diabetes  
 CC mellitus, lupus erythematosus, multiple sclerosis, myasthenia gravis,  
 CC Crohn's disease, rheumatoid arthritis, AIDS or ulcerative colitis),  
 CC cancer (including leukaemia, lymphoma, carcinomas of lung, penis, breast,  
 CC pancreas, colon, stomach, small intestine, brain or prostate), diseases  
 CC involving vesicle targeting, membrane transfer or fusion, or protein  
 CC processing, targeting or secretion, and conditions characterised by  
 CC abnormal levels of neurotransmitter production and release, lysosomal  
 CC storage diseases, chondrodysplasia, endocrine hypofunction and  
 CC hyperfunction.  
 XX Sequence 1175 BP; 387 A; 191 C; 244 G; 353 T; 0 other;  
 SQ Query Match 100.0%; Score 1175; DB 24; Length 1175;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-244;  
 Matches 1175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACTGTGATGAACACTTTTCCCGTGTGCTTTGAGTGCATCTTCAACAACCCCTAGGAG 60  
 DB 1 CACTGTGATGAACACTTTTCCCGTGTGCTTTGAGTGCATCTTCAACAACCCCTAGGAG 60  
 QY 61 GGTCTTGAAGCTTTTGAGATTAAACAATGGCAGGAAATCATCACTTTTAAAGTAATC 120  
 DB 61 GGTCTTGAAGCTTTTGAGATTAAACAATGGCAGGAAATCATCACTTTTAAAGTAATC 120  
 QY 121 TCCCTTGGAGATGGTGGAGTGGAGAGTTCATCTTATGAACAGATATGTAATAAAGT 180  
 DB 121 TCCCTTGGAGATGGTGGAGTGGAGAGTTCATCTTATGAACAGATATGTAATAAAGT 180  
 QY 181 TTGATACCCAGCTCTTCCATACAAATAGGTGTGGAATTTTAAATGAAGATTGGAGTGG 240  
 DB 181 TTGATACCCAGCTCTTCCATACAAATAGGTGTGGAATTTTAAATGAAGATTGGAGTGG 240  
 QY 241 ATGGACATTTTGTACCATGAGATTTGGACACCGCAGGTGAGGAGGATTCGGAAGCC 300  
 DB 241 ATGGACATTTTGTACCATGAGATTTGGACACCGCAGGTGAGGAGGATTCGGAAGCC 300  
 QY 301 TCAGACACCATTTTACAGAGTTCTGACTGTGCTGCTCTACTTTTATGTCGATGATT 360  
 DB 301 TCAGACACCATTTTACAGAGTTCTGACTGTGCTGCTCTACTTTTATGTCGATGATT 360  
 QY 361 CACAAAGCTTCCGAACCTTAAGTAATCTGGAAGAAAGAAATTCATATATTATGCAATGGA 420  
 DB 361 CACAAAGCTTCCGAACCTTAAGTAATCTGGAAGAAAGAAATTCATATATTATGCAATGGA 420  
 QY 421 AAGACCTGAGAGCTTCTCTTTTGTGATTTCTGGGTAAACAGATTGACATAGCGAAGCCG 480

DB 421 AAGACCTGAGAGCTTCTCTTTTGTGATTTCTGGGTAAACAGATTGACATAGCGAAGCCG 480  
 QY 481 AGGTGTCTACAGAAAGCCCAAGCTTTGGTGCAGGGAACAACGGGCACTATCTTTATTTTG 540  
 DB 481 AGGTGTCTACAGAAAGCCCAAGCTTTGGTGCAGGGAACAACGGGCACTATCTTTATTTTG 540  
 QY 541 AAACAAGTGAAGAGATGCGACAAATCTGCGACAGCCTTTGAGGAAGCGGTTGGAAGAG 600  
 DB 541 AAACAAGTGAAGAGATGCGACAAATCTGCGACAGCCTTTGAGGAAGCGGTTGGAAGAG 600  
 QY 601 TTTTGTCTACCGAGATAGTGCAGATCATTTGATTTACAGACAGACAGTCAATCTTCACC 660  
 DB 601 TTTTGTCTACCGAGATAGTGCAGATCATTTGATTTACAGACAGACAGTCAATCTTCACC 660  
 QY 661 GAAAGCCCAAGCCTAGCTCATCTTCTGTTGATTTGATTTGATTTGATTTGATTTTAAACA 720  
 DB 661 GAAAGCCCAAGCCTAGCTCATCTTCTGTTGATTTGATTTGATTTGATTTGATTTTAAACA 720  
 QY 721 ACTCACACATATACACAAATCAACATGGGATGGAGAGAGAAATAGCCCTTTGCAGCAG 780  
 DB 721 ACTCACACATATACACAAATCAACATGGGATGGAGAGAGAAATAGCCCTTTGCAGCAG 780  
 QY 781 TGTATCATCTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 840  
 DB 781 TGTATCATCTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 840  
 QY 841 GACATATCCACTCTTGGAGGAATATATTTACTCAATAATGGCAGCCTTACATTTTATAAAT 900  
 DB 841 GACATATCCACTCTTGGAGGAATATATTTACTCAATAATGGCAGCCTTACATTTTATAAAT 900  
 QY 901 GTAACAGTGTCTAATAACCGTTCTTTTAAATTTAAATTTAAATTTAAATTTAAATTT 960  
 DB 901 GTAACAGTGTCTAATAACCGTTCTTTTAAATTTAAATTTAAATTTAAATTTAAATTT 960  
 QY 961 GAAATGACCAAGACTTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1020  
 DB 961 GAAATGACCAAGACTTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1020  
 QY 1021 CTTGGATTTTCTCTGGGAAATGGAGAACTACTTTTATATGTTATGTTATGTTATGCAA 1080  
 DB 1021 CTTGGATTTTCTCTGGGAAATGGAGAACTACTTTTATATGTTATGTTATGTTATGCAA 1080  
 QY 1081 TTAGCATTTATTTCTTGGTTCAGGGAATACCTTTCTTAAGCAATATGTTAGATTTAA 1140  
 DB 1081 TTAGCATTTATTTCTTGGTTCAGGGAATACCTTTCTTAAGCAATATGTTAGATTTAA 1140  
 QY 1141 AGATTAAATCTAATGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1175  
 DB 1141 AGATTAAATCTAATGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1175

RESULT 3  
 ABZ35377  
 ID ABZ35377 standard; cDNA; 1372 BP.  
 XX ABZ35377;  
 AC ABZ35377;  
 XX 05-FEB-2003 (first entry)  
 DT 05-FEB-2003 (first entry)  
 DE Human gene expression profile polynucleotide SEQ ID NO 488.  
 KW Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;  
 KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;  
 KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;  
 KW gene expression; gene; ss.  
 OS Homo sapiens.  
 XX WO200274979-A2.  
 PN 26-SEP-2002.  
 XX 20-MAR-2002; 2002WO-US08456.



XX PR 20-MAR-2001; 2001US-276947P.  
XX PA (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC.  
XX PI Wan J, Wang Y;  
XX DR WPI; 2002-740862/80.  
XX PT New gene expression profile generated from primary, endothelial,  
PT epithelial, and muscle cell types, useful for identifying disease  
PT pathologies involving alterations of gene expression, e.g. cancer  
XX  
PS Disclosure; Page 631-632; 850pp; English.  
XX  
CC The invention relates to a gene expression profile comprising one or more  
CC genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type  
CC is a coronary artery endothelium, umbilical artery or vein endothelium,  
CC aortic endothelium, dermal microvascular endothelium, pulmonary artery  
CC endothelium, myometrium microvascular endothelium, keratinocyte  
CC epithelium, bronchial epithelium, mammary epithelium, prostate  
CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,  
CC small airway epithelium, renal epithelium, umbilical artery smooth  
CC muscle, neonatal dermal fibroblast, pulmonary artery smooth  
CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,  
CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,  
CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,  
CC osteoblasts or prostate stromal cell. The gene expression profile is used  
CC for determining the level of RNA expression for a sample, determining the  
CC phenotype of a cell and distinguishing cell types. The gene or a protein  
CC expression profile is useful in identifying disease pathologies  
CC involving alterations of gene expression. The assessment of expression  
CC profiles may provide meaningful information with respect to tumour type  
CC and stage, treatment methods, and prognosis. The gene or protein  
CC expression profile may also be used for creating microarrays. The  
CC microarray is useful for genetic and physical mapping of genomes, DNA  
CC sequencing, genetic or medical diagnosis, genotyping of organisms,  
CC confirming cell or tissue identifications and in identifying promising  
CC antibiotics, antiviral or antifungal agents.  
XX  
SQ Sequence 1372 BP; 404 A; 261 C; 306 G; 397 T; 4 other;  
Query Match 98.7%; Score 1160.2; DB 24; Length 1372;  
Best Local Similarity 99.2%; Pred No. 9.5e-241;  
Matches 1165; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 CACTGTGATGAACACTTTCCCGTGTGCTTTGAGTGCATCTTCTCAACAACCCCTAGGAG 60  
Db 190 CACTGTGATGAACACTTTCCCGTGTGCTTTGAGTGCATCTTCTCAACAACCCCTAGGAG 249  
QY 61 GGTTCCTTGAAGCTTTTGAGATTAACTGGCAGGAAATCATCATCTTTTAAAGTAATTC 120  
Db 250 GGTTCCTTGAAGCTTTTGAGATTAACTGGCAGGAAATCATCATCTTTTAAAGTAATTC 309  
QY 121 TCCTTGGAGATGGTGGAGTTGGGAAGAGTTCACTTATGAACAGATATCTAATAAAGT 180  
Db 310 TCCTTGGAGATGGTGGAGTTGGGAAGAGTTCACTTATGAACAGATATCTAATAAAGT 369  
QY 181 TTGATACCCAGCTCTTCCATACAAATAGGTGGGAATTTTAAATAAGATTGGAAGTGG 240  
Db 370 TTGATACCCAGCTCTTCCATACAAATAGGTGGGAATTTTAAATAAGATTGGAAGTGG 429  
QY 241 ATGGACATTTTCTTACCATGTCAGATTGGGACACGGCAGGTCAGGAGGATTCGGAAGCC 300  
Db 430 ATGGACATTTTCTTACCATGTCAGATTGGGACACGGCAGGTCAGGAGGATTCGGAAGCC 489  
QY 301 TGAGGACACCATTTTACAGAGGTTCTGACTGCTGCTGCTTACTTTTAAAGTTCGATGATT 360  
Db 490 TGAGGACACCATTTTACAGAGGTTCTGACTGCTGCTGCTTACTTTTAAAGTTCGATGATT 549  
QY 361 CACAAAGCTTCCAGAACTTAACTAAGTGAAGAAAGATTTCATATATTATGCAGATGGA 420  
Db 550 CACAAAGCTTCCAGAACTTAACTAAGTGAAGAAAGATTTCATATATTATGCAGATGGA 609

QY 421 AAGAGCCTGAGAGCTTTTCCTTTTGTGATTCTGGGTAAACAAGATTGACATGAAGCGAACGGC 480  
Db 610 AAGAGCCTGAGAGCTTTTCCTTTTGTGATTCTGGGTAAACAAGATTGACATGAAGCGAACGGC 669  
QY 481 AGGTGCTCTACAGAAGAGCCCAAGCTTTGGTGCAGGAGCAACGGCGACTATCCCTTATTTTG 540  
Db 570 AGGTGCTCTACAGAAGAGCCCAAGCTTTGGTGCAGGAGCAACGGCGACTATCCCTTATTTTG 729  
QY 541 AAACAAGTGCAAAAGATGCCCAAAATGTGGCAGCAGCCTTTGAGGAAGCGTTTCGAAGAG 600  
Db 730 AAACAAGTGCAAAAGATGCCCAAAATGTGGCAGCAGCCTTTGAGGAAGCGTTTCGAAGAG 789  
QY 601 TTCTTGCTACCGAGGATAGTGCAGATCATTTGATTTCAGACAGACACAGTCAATCTTCACC 660  
Db 790 TTCTTGCTACCGAGGATAGTGCAGATCATTTGATTTCAGACAGACACAGTCAATCTTCACC 849  
QY 661 GAAAGCCCAAGCCTAGCTCATCTTGTGTTGATTGTTAGATTGTTGATGCTTCTTAACCA 720  
Db 850 GAAAGCCCAAGCCTAGCTCATCTTGTGTTGATTGTTAGATTGTTGATGCTTCTTAACCA 909  
QY 721 ACTCACACATATACACAAATCAACATGGGATGGGAGAGAGATTTAGCGTTTCGAGCAG 780  
Db 910 ACTCACACATATACACAAATCAACATGGGATGGGAGAGAGATTTAGCGTTTCGAGCAG 969  
QY 781 TGTATCATCTACTATAATAAATTAACCTAATGTTGCTGCTTCTTCAATTTAGTTGGTGGAGAG 840  
Db 970 TGTATCATCTACTATAATAAATTAACCTAATGTTGCTGCTTCTTCAATTTAGTTGGTGGAGAG 1029  
QY 841 GACACATCCACTCTTGGAGGATATATTTACTCAATAATGGCAGCCTTACATTTATAAAT 900  
Db 1030 GACACATCCACTCTTGGAGGATATATTTACTCAATAATGGCAGCCTTACATTTATAAAT 1089  
QY 901 GTAAACAGTTGTCTAAATAACGTTCTTTTAAATTTAAATATCTAAGTTTCAGAGCTAATAAT 960  
Db 1090 GTAAACAGTTGTCTAAATAACGTTCTTTTAAATTTAAATATCTAAGTTTCAGAGCTAATAAT 1149  
QY 961 GAAATGACCAAGACTTTTAAATATAATAAATAAAGAACTTGACTTATTTAGAGTTATA 1020  
Db 1150 GAAATGACCAAGACTTTTAAATATAATAAATAAAGAACTTGACTTATTTAGAGTTATA 1209  
QY 1021 CTTGGATTTTTCCTGGGAAATGGAGAACTACTTTTATATATGTTGTTATGCTTATGCAA 1080  
Db 1210 CTTGGATTTTTCCTGGGAAATGGAGAACTACTTTTATATGTTGTTATGCTTATGCAA 1269  
QY 1081 TTAGCATTTGATTCTTTGGTTCAGGGAAATACCTTTCTAAGCAATAATGTTAGATATTAA 1140  
Db 1270 TTAGCATTTGATTCTTTGGTTCAGGGAAATACCTTTCTAAGCAATAATGTTAGATATTAA 1329  
QY 1141 AGATTAAATCTAATGTTAAATAAATAAATAAATAA 1174  
Db 1330 AGATTAAATCTAATGTTAAATAAATAAATAAANA 1363  
RESULT 4  
ABA45112  
ID ABA45112 standard; DNA; 1969 BP.  
XX ABA45112;  
AC ABA45112;  
XX  
DT 01-FEB-2002 (first entry)  
XX  
DE Human breast cell single exon nucleic acid probe #3807.  
XX  
KW Human; microarray; single exon probe; gene expression; breast;  
KW disease; cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157271-A2.  
XX  
XX 09-AUG-2001.  
PD  
XX

PF 30-JAN-2001; 2001WO-US00662.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-496933/54.  
XX  
DR New spatially-addressable set of single exon nucleic acid probes,  
XX useful for measuring gene expression in sample derived from human  
PT breast, comprises number of single exon nucleic acid probes -  
XX  
PS Claim 1; SEQ ID NO 3807; 327pp + sequence listing; English.  
XX  
CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and Bt 474 cells. The method involves contacting  
CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1969 BP; 624 A; 330 C; 408 G; 607 T; 0 other;

Query Match 93.5%; Score 1100; DB 22; Length 1969;  
Best Local Similarity 100.0%; Pred. No. 9.8e-228;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

59 AGGGTCTTGAAGCTTTGAGATTAACAATGGCAGGAAATCATCACTTTTAAAGTAAT 118  
311 AGGGTCTTGAAGCTTTGAGATTAACAATGGCAGGAAATCATCACTTTTAAAGTAAT 370

119 TCTCCTTGGAGATGGTGGAGTGGGAGAGTTCACCTATGACAGATATGTAACATAATA 178  
371 TCTCCTTGGAGATGGTGGAGTGGGAGAGTTCACCTATGACAGATATGTAACATAATA 430

179 GTTTGATACCCAGCTTCCATACATAGTGTGGAAATTTTAAATAAGATTTGGAAGT 238  
431 GTTTGATACCCAGCTTCCATACATAGTGTGGAAATTTTAAATAAGATTTGGAAGT 490

239 GGATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTTCAGGACGATTCGGAAG 298  
491 GGATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTTCAGGACGATTCGGAAG 550

299 CTTGAGGACACATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAGTGTGATGA 358  
551 CTTGAGGACACATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAGTGTGATGA 610

359 TTCACAAAGCTTCCAGAACTTAACTGGAAGAAAGAAATTCATATATATGCGAGATGT 418  
611 TTCACAAAGCTTCCAGAACTTAACTGGAAGAAAGAAATTCATATATATGCGAGATGT 670

419 GAAAGAGCTGAGAGCTTTCCTTTTGTGATCTGGGTAAACAGATTCAGGAGCAACG 478

Db 671 GAAAGAGCTGAGAGCTTTCCTTTTGTGATCTGGGTAAACAGATTCAGATAACGGAACG 730  
QY 479 GCAGGTGTCTACAGAAAGCCCAAGCTTGTGTGAGGGAACAACGGCACTATCCTTATTT 538  
Db 731 GCAGGTGTCTACAGAAAGCCCAAGCTTGTGTGAGGGAACAACGGCACTATCCTTATTT 790  
QY 539 TGAACAAAGTGCACAAAGATGCCAATATGTGGCAGCAGCCCTTTGAGGAAGCGGTTCGAAG 598  
Db 791 TGAACAAAGTGCACAAAGATGCCAATATGTGGCAGCAGCCCTTTGAGGAAGCGGTTCGAAG 850  
QY 599 AGTTCTTGTCTACCGAGGATAGGTGAGATCAATTTGATTCAGACACACAGCAATCTTCA 658  
Db 851 AGTTCTTGTCTACCGAGGATAGGTGAGATCAATTTGATTCAGACACACAGCAATCTTCA 910  
QY 659 CCGAAGCCCAAGCTGAGCTCATCTTGTGTGATTTGATTTGATGATTTGATGATTTTAAAC 718  
Db 911 CCGAAGCCCAAGCTGAGCTCATCTTGTGTGATTTGATTTGATGATTTGATGATTTTAAAC 970  
QY 719 CAATCTCACATATACACAAATCAACATGGGATGGAGAAAGAAATAGCGTTTGCAGC 778  
Db 971 CAATCTCACATATACACAAATCAACATGGGATGGAGAAAGAAATAGCGTTTGCAGC 1030  
QY 779 AGTGTATCATCTACTATATAAAATTAACATATGTTGCTGCTTCAATAGTTGGTGGAGAA 838  
Db 1031 AGTGTATCATCTACTATATAAAATTAACATATGTTGCTGCTTCAATAGTTGGTGGAGAA 1090  
QY 839 GGGACACATCCACCTTTGGAGGATATATTTACTCAATATGCGACCTTACATTTATAAA 898  
Db 1091 GGGACACATCCACCTTTGGAGGATATATTTACTCAATATGCGACCTTACATTTATAAA 1150  
QY 899 TTGTAAACAGTTGTCTAATAACGTTTCTTTAATTTAATAATATGTAAGTTGCGAGCTAATAA 958  
Db 1151 TTGTAAACAGTTGTCTAATAACGTTTCTTTAATTTAATAATATGTAAGTTGCGAGCTAATAA 1210  
QY 959 ATGAAATGACCAAGCACTTTAATTTAATAATAAAATAAGAACTTACATTTCTAGAAGTTA 1018  
Db 1211 ATGAAATGACCAAGCACTTTAATTTAATAATAAAATAAGAACTTACATTTCTAGAAGTTA 1270  
QY 1019 TACTTGGATTTTTCCTGGGAAATGGAGAACTACTTTTATATGTTATGTTTATGCTATGCT 1078  
Db 1271 TACTTGGATTTTTCCTGGGAAATGGAGAACTACTTTTATATGTTATGTTTATGCTATGCT 1330  
QY 1079 AATTAGCAATGTATCTTGTTCAGGAAATACCTTCTTAAGCAATATGTTAGATATT 1138  
Db 1331 AATTAGCAATGTATCTTGTTCAGGAAATACCTTCTTAAGCAATATGTTAGATATT 1390  
QY 1139 AAAGATTAATAATCTAATGTA 1158  
Db 1391 AAAGATTAATAATCTAATGTA 1410

RESULT 5  
AAK03817  
ID AAK03817 standard; DNA; 1969 BP.  
XX  
AC AAK03817;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe SEQ ID NO: 3808.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00667.  
XX

PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-483446/52.  
 XX  
 XX Single exon nucleic acid probes for analyzing gene expression in human  
 XX brains -  
 XX  
 XX Example 4; SEQ ID NO: 3808; 650pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention.  
 XX  
 XX Sequence 1969 BP; 624 A; 330 C; 408 G; 607 T; 0 other;  
 SQ  
 Query Match 93.6%; Score 1100; DB 228; Length 1969;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-228;  
 Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 59 AGGGTCTTGAAGCTTTGAGATTAACAATGGCAGGAAATCATCATTAAAGTAAT 118  
 Db 311 AGGGTCTTGAAGCTTTGAGATTAACAATGGCAGGAAATCATCATTAAAGTAAT 370  
 Qy 119 TCTCCTTGGAGTGGGAGTTGGGAGAGTTCACTTATGAACAGATATGTAACATAA 178  
 Db 371 TCTCCTTGGAGTGGGAGTTGGGAGAGTTCACTTATGAACAGATATGTAACATAA 430  
 Qy 179 GTTTGATACCCAGCTCTCCATACATAGTGTGGAATTTTAAATAAGATTTGGAAGT 238  
 Db 431 GTTTGATACCCAGCTCTCCATACATAGTGTGGAATTTTAAATAAGATTTGGAAGT 490  
 Qy 239 GGATGACATTTTGTACATGACAGATTTGGGACACGCGAGTCCAGGCGATTCGGAAG 298  
 Db 491 GGATGACATTTTGTACATGACAGATTTGGGACACGCGAGTCCAGGCGATTCGGAAG 550  
 Qy 299 CCTGAGGACACCATTTTACAGAGGTTCTGACTGCTGCTGCTTACTTTTAGTGTCGATGA 358  
 Db 551 CCTGAGGACACCATTTTACAGAGGTTCTGACTGCTGCTGCTTACTTTTAGTGTCGATGA 610  
 Qy 359 TTCAAAAGCTTCCAGAACTTAAGTAACTGGGAGAAAGAAATTCATATATATGCAAGATGT 418  
 Db 611 TTCAAAAGCTTCCAGAACTTAAGTAACTGGGAGAAAGAAATTCATATATATGCAAGATGT 670  
 Qy 419 GAAAGAGCTGAGAGCTTTCTTTTGTGATCTGGTAACTGATGACATAGCGAAG 478  
 Db 671 GAAAGAGCTGAGAGCTTTCTTTTGTGATCTGGTAACTGATGACATAGCGAAG 730  
 Qy 479 GCAGTGTCTACAGAAAGCCCAAGCTTTGGTGCAGGAGCAACGGCGACTATCTCTATT 538  
 Db 731 GCAGTGTCTACAGAAAGCCCAAGCTTTGGTGCAGGAGCAACGGCGACTATCTCTATT 790  
 Qy 539 TGAACAAGTGCAGAGATGCCAAGTGTGGCAGCGCTTTGAGGAGCGGTTGGAAG 598  
 Db 791 TGAACAAGTGCAGAGATGCCAAGTGTGGCAGCGCTTTGAGGAGCGGTTGGAAG 850  
 Qy 599 AGTTCTTGTACCGAGGATAGTTCAGATCATTTGATTTCAGACAGACAGTCAATCTTCA 658  
 Db 851 AGTTCTTGTACCGAGGATAGTTCAGATCATTTGATTTCAGACAGACAGTCAATCTTCA 910

Qy 659 CCGAAGCCCAAGCCTAGCTCATCTTCTGTTGATTTGTTAGATTTGTTGATGCTTCTTAAAC 718  
 Db 911 CCGAAGCCCAAGCCTAGCTCATCTTCTGTTGATTTGTTAGATTTGTTGATGCTTCTTAAAC 970  
 Qy 719 CAACCTCACATATACACAAAATCAACATGGGATGGAGAGAGAAATAGCGTTTGCAGC 778  
 Db 971 CAACCTCACATATACACAAAATCAACATGGGATGGAGAGAGAAATAGCGTTTGCAGC 1030  
 Qy 779 AGTGTATCATCTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 838  
 Db 1031 AGTGTATCATCTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1090  
 Qy 839 GGGACACATCCACTCTTGGAGGAATATATTTACTCTCAATAATGCACCTTACATTTATAA 898  
 Db 1091 GGGACACATCCACTCTTGGAGGAATATATTTACTCTCAATAATGCACCTTACATTTATAA 1150  
 Qy 899 TTGTAAACAGTTGCTAATAAACGTTTCTTTAAATAAATAAATAAATAAATAAATAAATAA 958  
 Db 1151 TTGTAAACAGTTGCTAATAAACGTTTCTTTAAATAAATAAATAAATAAATAAATAAATAA 1210  
 Qy 959 ATGAAATGACCAAGACTTTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1018  
 Db 1211 ATGAAATGACCAAGACTTTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1270  
 Qy 1019 TACTTGGATTTTTCCTGGGAAAATGGAGAACTACTTTTATATGTAATGTTTATGCT 1078  
 Db 1271 TACTTGGATTTTTCCTGGGAAAATGGAGAACTACTTTTATATGTAATGTTTATGCT 1330  
 Qy 1079 AATTAGCATTTGATCTTCTGCTCAGGGAATACTTTCTTAAAGCAATATGTTAGATATT 1138  
 Db 1331 AATTAGCATTTGATCTTCTGCTCAGGGAATACTTTCTTAAAGCAATATGTTAGATATT 1390  
 Qy 1139 AAGATTAAATAATCTAATGTA 1158  
 Db 1391 AAGATTAAATAATCTAATGTA 1410  
 RESULT 6  
 AA113885  
 ID AA113885 standard; DNA; 1969 BP.  
 XX  
 AC AA113885;  
 XX  
 XX  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Probe #3818 for gene expression analysis in human cervical cell sample.  
 XX  
 KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer; ss.  
 OS Homo sapiens.  
 XX  
 FN WO200157278-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00670.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488901/53.  
 XX  
 XX Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human cervical epithelial cells -

Claim 25; SEQ ID No 3918; 487pp; English.

The present invention relates to human single exon nucleic acid probes (SENPs). The present sequence is one such probe. The SENPs are derived from human Hela cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pet\\_sequences](http://ftp.wipo.int/pub/published_pet_sequences).

Sequence 1969 BP; 624 A; 330 C; 408 G; 607 T; 0 other;

Query Match 93.6%; Score 1100; DB 22; Length 1969;

Best Local Similarity 100.0%; Pred. No. 9.8e-228;

Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	59	AGGGTTCTTGAGCTTTTCGAGTTACATGCAATGCGGAAATCATCACCTTTTAAAGTAA	118
DB	311	AGGGTTCTTGAGCTTTTCGAGTTACATGCAATGCGGAAATCATCACCTTTTAAAGTAA	370
QY	119	TCTCCTTGGAGATGGTGGAGTTGGGAGAGTTTCACTTATGAACAGATATGTAACATAA	178
DB	371	TCTCCTTGGAGATGGTGGAGTTGGGAGAGTTTCACTTATGAACAGATATGTAACATAA	430
QY	179	GTTTGATACCCAGCTTCCATACATAGTGTGGAAATTTTAAATAAGATTGGAAGT	238
DB	431	GTTTGATACCCAGCTTCCATACATAGTGTGGAAATTTTAAATAAGATTGGAAGT	490
QY	239	GGATGACATTTTGTACCATGAGATTTGGGACACGGCAGGTGAGGACGATTCGGAAG	298
DB	491	GGATGACATTTTGTACCATGAGATTTGGGACACGGCAGGTGAGGACGATTCGGAAG	550
QY	299	CCTGAGGACACATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTGTGTCGATGA	358
DB	551	CCTGAGGACACATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTGTGTCGATGA	610
QY	359	TTTCAAAAGCTTCCAGAACTTAACTGGAAGAAAGAAATTCATATATTATGCAAGAT	418
DB	611	TTTCAAAAGCTTCCAGAACTTAACTGGAAGAAAGAAATTCATATATTATGCAAGAT	670
QY	419	GAAAGAGCTGAGAGCTTTCTCTTTTGTGATTTCTGGTAAACAAGATTGACATAAGCGA	478
DB	671	GAAAGAGCTGAGAGCTTTCTCTTTTGTGATTTCTGGTAAACAAGATTGACATAAGCGA	730
QY	479	GCAGGTGCTTACAGAGAGCCAGCTTTGTCAGGACCAACGGGACGATTCCTTATTT	538
DB	731	GCAGGTGCTTACAGAGAGCCAGCTTTGTCAGGACCAACGGGACGATTCCTTATTT	790
QY	539	TGAACAAGTGCAGAAAGATCCCAAAATGTCGACAGCTTTTGGAGAGCGGTTTCGAAG	598
DB	791	TGAACAAGTGCAGAAAGATCCCAAAATGTCGACAGCTTTTGGAGAGCGGTTTCGAAG	850
QY	599	AGTTCTTGCTACGAGGATAGTCAGATCATTTGATTCAGACAGACAGTCAATCTTCA	658
DB	851	AGTTCTTGCTACGAGGATAGTCAGATCATTTGATTCAGACAGACAGTCAATCTTCA	910
QY	659	CCGAAAGCCCAAGCTTAGCTCATCTTCTGTTGATTTGTAGATTGTTGATGATTCATAC	718
DB	911	CCGAAAGCCCAAGCTTAGCTCATCTTCTGTTGATTTGTAGATTGTTGATGATTCATAC	970
QY	719	CAACTCACATATACAAAAATCAACATGCGGATGGAGAGAAATTAGCGTTTCGAGC	778
DB	971	CAACTCACATATACAAAAATCAACATGCGGATGGAGAGAAATTAGCGTTTCGAGC	1030
QY	779	AGTGATCATCTACTAATAAATAAATAAAGTTCCTGCTTCATTTAGTTGGTGGAGAA	838
DB	1031	AGTGATCATCTACTAATAAATAAATAAAGTTCCTGCTTCATTTAGTTGGTGGAGAA	1090

QY	839	GGGACATCCACTCTTGGAGGAATATATTACTCAATAATGGCACCTTACATTATAAA	898
DB	1091	GGGACATCCACTCTTGGAGGAATATATTACTCAATAATGGCACCTTACATTATAAA	1150
QY	899	TTGTAACAGATTGCTAATAACGTTTCTTAATTTAAATATGTAAGTTGCAGAGCTATAA	958
DB	1151	TTGTAACAGATTGCTAATAACGTTTCTTAATTTAAATATGTAAGTTGCAGAGCTATAA	1210
QY	959	ATGAATGACCAAGACTTTAATTAATAAATAAAGAACTTGACTATTCTAGAGTTA	1018
DB	1211	ATGAATGACCAAGACTTTAATTAATAAATAAAGAACTTGACTATTCTAGAGTTA	1270
QY	1019	TACTTGATTTTTCCTCGGAAATGAGAACTACTTTTATATGCTATGTTTATGTC	1078
DB	1271	TACTTGATTTTTCCTCGGAAATGAGAACTACTTTTATATGCTATGTTTATGTC	1330
QY	1079	AATTAGCATTTTCTTGGTTCCAGGAAATCTTTCTTAAAGCAATAATGTTAGATATT	1138
DB	1331	AATTAGCATTTTCTTGGTTCCAGGAAATCTTTCTTAAAGCAATAATGTTAGATATT	1390
QY	1139	AAAGATTAAATCTAATGTA	1158
DB	1391	AAAGATTAAATCTAATGTA	1410

RESULT 7

AAI03747

ID AAI03747 standard; DNA; 1969 BP.

XX AAI03747;

DT 09-OCT-2001 (first entry)

Probe #3738 used to measure gene expression in human breast sample.

Probe; human; breast disease; breast cancer; development disorder; ss;

inflammatory disease; proliferative breast disease; non-carcinoma tumour.

Homo sapiens.

WO200157270-A2.

09-AUG-2001.

29-JAN-2001; 2001WO-US000661.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0832366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-476286/51.

Novel single exon nucleic acid probe used to measuring gene expression in a human breast -

Claim 25; SEQ ID No 3738; 322pp; English.

The present invention relates to novel single exon nucleic acid probes.

The present sequence is one such probe. The probes are useful for

measuring human gene expression in a human breast sample, where the probe

hybridises at high stringency to a nucleic acid expressed in the human

breast. The probes are useful for predicting, diagnosing, grading,

staging, monitoring and prognosing diseases of the human breast,

particularly those diseases with polygenic aetiology. The diseases

include: breast cancer, disorders of development, inflammatory diseases

CC of the breast, fibrocystic changes, proliferative breast disease and  
 CC non-carcinoma tumours.  
 CC Note: the sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1969 BP; 624 A; 330 C; 408 G; 607 T; 0 other;

Query Match 93.6%; Score 1100; DB 22; Length 1969;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-228; Mismatches 0; Indels 0; Gaps 0;  
 Matches 1100; Conservative 0;

QY 59 AGGGTTCTTGAAGCTTTTGGAGTAAACAATGCGAGAAATCATCACTTTTAAAGTAAT 118  
 DB 311 AGGGTTCTTGAAGCTTTTGGAGTAAACAATGCGAGAAATCATCACTTTTAAAGTAAT 370

QY 119 TCTCTTGGAGATGCTGGAGTTGGAGAGTTCATTATGAACAGATATGTAACATAAA 178  
 DB 371 TCTCTTGGAGATGCTGGAGTTGGAGAGTTCATTATGAACAGATATGTAACATAAA 430

QY 179 GTTTCATACCCAGCTCTTCCATACAAATAGTGTGGAATTTTAAATAAAGATTTGGAAGT 238  
 DB 431 GTTTCATACCCAGCTCTTCCATACAAATAGTGTGGAATTTTAAATAAAGATTTGGAAGT 490

QY 239 GGATGGAATTTGTTTACATGACATTTGGGACACGCGAGGTGAGGAGCATTCGGAAG 298  
 DB 491 GGATGGAATTTGTTTACATGACATTTGGGACACGCGAGGTGAGGAGCATTCGGAAG 550

QY 299 CTTGAGGACACATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTGTAGTGTGATGA 358  
 DB 551 CTTGAGGACACATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTGTAGTGTGATGA 610

QY 359 TTCAACAAGCTTCCAGAACTTAAAGTAATGGAAGAAAGATTCATATATGAGATGT 418  
 DB 611 TTCAACAAGCTTCCAGAACTTAAAGTAATGGAAGAAAGATTCATATATGAGATGT 670

QY 419 GAAAGAGCTGAGAGCTTCTTTTGTGATCTGGTAAACAAGATTCACATAAGCGAAG 478  
 DB 671 GAAAGAGCTGAGAGCTTCTTTTGTGATCTGGTAAACAAGATTCACATAAGCGAAG 730

QY 479 GCAGGTGTCTACAGAAAGCCCAAGCTTGTGTCAGGAGCAACCGCACTATCTCTTATTT 538  
 DB 731 GCAGGTGTCTACAGAAAGCCCAAGCTTGTGTCAGGAGCAACCGCACTATCTCTTATTT 790

QY 539 TGAACAAGTGCAGAAAGATGCCAACAATGTGGCAGCAGCCCTTTCAGAGAGCGTTTCAAG 598  
 DB 791 TGAACAAGTGCAGAAAGATGCCAACAATGTGGCAGCAGCCCTTTCAGAGAGCGTTTCAAG 850

QY 599 AGTTCTTGTCTACCGAGGATAGTTCAGATCATTTGATTCAGACACACAGTCAATCTTCA 658  
 DB 851 AGTTCTTGTCTACCGAGGATAGTTCAGATCATTTGATTCAGACACACAGTCAATCTTCA 910

QY 659 CCGAAAGCCCAAGCTTAGTCACTTCTGCTGTTGATTTAGATTTGATGATGATCTTCAAC 718  
 DB 911 CCGAAAGCCCAAGCTTAGTCACTTCTGCTGTTGATTTAGATTTGATGATCTTCAAC 970

QY 719 CAACTCACAATATACACAATAACAATGCGATGGGATGGAGAGAAATTTAGCGTTTGCAGC 778  
 DB 971 CAACTCACAATATACACAATAACAATGCGATGGGATGGAGAGAAATTTAGCGTTTGCAGC 1030

QY 779 AGTGTATCATCTACTAATAAATAAATAAATTAATTTGCTGCTTCAATTTAGTTGGTGGGAA 838  
 DB 1031 AGTGTATCATCTACTAATAAATAAATAAATTAATTTGCTGCTTCAATTTAGTTGGTGGGAA 1090

QY 839 GGGACACATCCACTCTTGGAGAAATATATTTACTCAATAAATGGCAGCTTACATTTATAA 898  
 DB 1091 GGGACACATCCACTCTTGGAGAAATATATTTACTCAATAAATGGCAGCTTACATTTATAA 1150

QY 899 TTGTAACAGTGTCTAATAAAGCTTTCTTTTAAATATATGTAAGTTTCAGAGCTAATAA 958  
 DB 1151 TTGTAACAGTGTCTAATAAAGCTTTCTTTTAAATATATGTAAGTTTCAGAGCTAATAA 1210

QY 959 ATGAATGACCAAGACTTTTATATATAATAAATAAAGAACTTGACTTATCTTAGAAGTTA 1018

DB 1211 ATGAATGACCAAGACTTTAATTAATAAATAAATAAAGAACTTGACTTATCTAGAAGTTA 1270

QY 1019 TACTTCGATTTTCTCTGGAAAAATGGAGAACTATCTTTTATATGATGTTTATGTC 1078

DB 1271 TACTTCGATTTTCTCTGGAAAAATGGAGAACTATCTTTTATATGATGTTTATGTC 1330

QY 1079 AATTAGCATTTGTTCTGTTTCAGGAAATACTTTCTTAAAGCAATATGTTAGATATT 1138

DB 1331 AATTAGCATTTGTTCTGTTTCAGGAAATACTTTCTTAAAGCAATATGTTAGATATT 1390

QY 1139 AAAGATTAATAATCTAATGTA 1158

DB 1391 AAAGATTAATAATCTAATGTA 1410

RESULT 8  
 ABS03847  
 ID ABS03847 standard; DNA; 1969 BP.  
 XX  
 AC ABS03847;  
 XX  
 DT 19-AUG-2002 (first entry)  
 XX  
 DE Human genome-derived single exon probe from lung SEQ ID No 3838.  
 XX  
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PP 30-JAN-2001; 2001WO-US00665.  
 XX  
 PR 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 XX WPI; 2002-114183/15.  
 DR  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX  
 PS Claim 1; SEQ ID No 3838; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung

CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic,  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and lysine membrane disease. The present sequence is a single exon  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1969 BP; 624 A; 330 C; 408 G; 607 T; 0 other;

Query Match 93.6%; Score 1100; DB 24; Length 1969;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-228;  
 Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 AGGGTCTTGAAGCTTTTGAGATTACAAATGGCAGGAAATCATCATCTTTTAAAGTAAT 118  
 DB 311 AGGGTCTTGAAGCTTTTGAGATTACAAATGGCAGGAAATCATCATCTTTTAAAGTAAT 370  
 QY 119 TCTCCTTGGAGATGGTGGAGTTCGGGAGAGTTCACATTATGAACAGATATGTAACATAAA 178  
 DB 371 TCTCCTTGGAGATGGTGGAGTTCGGGAGAGTTCACATTATGAACAGATATGTAACATAAA 430  
 QY 179 GTTTGATACCCAGCTCTTCCATACAAATAGGTGTGGAATTTTAAATAAGATTGGAAT 238  
 DB 431 GTTTGATACCCAGCTCTTCCATACAAATAGGTGTGGAATTTTAAATAAGATTGGAAT 490  
 QY 239 GGATGGACATTTTGTACCATGCAGATTTGGACACGGCAGGTTCAGGACGATTCGGAAG 298  
 DB 491 GGATGGACATTTTGTACCATGCAGATTTGGACACGGCAGGTTCAGGACGATTCGGAAG 550  
 QY 299 CTTGAGGACACCACTTTTACAGAGTTCTGACTGCTGCTTCTTCTTCTTCTTCTTCTTCTT 358  
 DB 551 CTTGAGGACACCACTTTTACAGAGTTCTGACTGCTGCTTCTTCTTCTTCTTCTTCTTCTT 610  
 QY 359 TTCCAGAGCTCCAGACTTAAAGTAACTGAGAGAGATTCATATATATGTCAGATGT 418  
 DB 611 TTCCAGAGCTCCAGACTTAAAGTAACTGAGAGAGATTCATATATATGTCAGATGT 670  
 QY 419 GAAAGAGCTGAGAGCTTTCTTTTCTGATTTCTGGGTAAACAAGATTGACATAAGCGAAG 478  
 DB 671 GAAAGAGCTGAGAGCTTTCTTTTCTGATTTCTGGGTAAACAAGATTGACATAAGCGAAG 730  
 QY 479 GCAGGTGTCTACAGAGAGAGCCCAAGCTTTGGTGACGGGCAACAGCGGCACTATCTTATTT 538  
 DB 731 GCAGGTGTCTACAGAGAGAGCCCAAGCTTTGGTGACGGGCAACAGCGGCACTATCTTATTT 790  
 QY 539 TGAACAAGTGAAGAGAGTCCCAAAATGTGGCAGCAGCTTTTGGAGAGCGGTTTCGAAG 598  
 DB 791 TGAACAAGTGAAGAGAGTCCCAAAATGTGGCAGCAGCTTTTGGAGAGCGGTTTCGAAG 850

QY 599 AGTTCTTGCTACCGAGGATAGGTCAGATCATTTGATTTCAGACAGACACAGCTCAATCTTCA 658  
 DB 851 AGTTCTTGCTACCGAGGATAGGTCAGATCATTTGATTTCAGACAGACACAGCTCAATCTTCA 910  
 QY 659 CCGAAAGCCCAAGCCTAGCTCATCTTCTGCTGTTAGTTTGTAGTTGTTGATGCTTCTTAAAC 718  
 DB 911 CCGAAAGCCCAAGCCTAGCTCATCTTCTGCTGTTAGTTTGTAGTTGTTGATGCTTCTTAAAC 970  
 QY 719 CAATCTCACATATACACAAATCAACAAATCAACAAATCAACAAATCAACAAATCAACAAATCA 778  
 DB 971 CAATCTCACATATACACAAATCAACAAATCAACAAATCAACAAATCAACAAATCAACAAATCA 1030  
 QY 779 AGTGTATCATCTACTATAATAAAATTAATACTAATGTGCTTCTTCAATAGTTGGTGGAGAA 838  
 DB 1031 AGTGTATCATCTACTATAATAAAATTAATACTAATGTGCTTCTTCAATAGTTGGTGGAGAA 1090  
 QY 839 GGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTATAAA 898  
 DB 1091 GGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTATAAA 1150  
 QY 899 TTGTAAACAGTTGCTCTAATAACGTTTCTTAAATTAATAATGTAAGTTGCAGAGCTAATAA 958  
 DB 1151 TTGTAAACAGTTGCTCTAATAACGTTTCTTAAATTAATAATGTAAGTTGCAGAGCTAATAA 1210  
 QY 959 ATGAAATGACCAAGACTTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATA 1018  
 DB 1211 ATGAAATGACCAAGACTTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATA 1270  
 QY 1019 TACTTGGATTTTCTCTGGGAAATGCGAAGTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1078  
 DB 1271 TACTTGGATTTTCTCTGGGAAATGCGAAGTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1330  
 QY 1079 AATTAGCATTTGATTTCTTGGTTTCAGGGAAATACCTTCTCTAAAGCAATAATGTTAGATATT 1138  
 DB 1331 AATTAGCATTTGATTTCTTGGTTTCAGGGAAATACCTTCTCTAAAGCAATAATGTTAGATATT 1390  
 QY 1139 AAGATTAAATCTAATGTA 1158  
 DB 1391 AAGATTAAATCTAATGTA 1410  
 RESULT 9  
 AAL02805  
 ID AAL02805 standard; DNA; 6058 BP.  
 XX AC AAL02805;  
 XX DT 21-NOV-2001 (first entry)  
 XX XX  
 DE Human reproductive system related antigen DNA SEQ ID NO: 5493.  
 XX KW Human; reproductive system related antigen; reproductive system disorder;  
 XX KW cancer; gene therapy; ds.  
 XX OS Homo sapiens.  
 XX PN WO200155320-A2.  
 XX PD 02-AUG-2001.  
 XX PF 17-JAN-2001; 2001WO-US01339.  
 XX PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 07-MAY-2000; 2000US-020515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216547.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 03-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen  
XX is used in preventing, treating or ameliorating a medical condition -  
XX  
XX Disclosure; SEQ ID NO 5493; 1297pp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
XX number of human reproductive system related antigens. These can be used  
XX in the prevention and treatment of reproductive system disorders,  
XX including cancer. The present sequence is a genomic sequence encoding a  
XX protein of the invention.  
XX  
XX Sequence 6058 BP; 1764 A; 1068 C; 1313 G; 1913 T; 0 other;  
SQ

Query Match

93.8%; Score 1100; DB 22; Length 6058;

Best Local Similarity 100.0%; Pred. No. 1.2e-227; Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 AGGTTCTTGAAGCTTTGAGATTAAACAATGGCAGGAAATCATCTTTTAAAGTAAT 118  
 Db 4905 AGGTTCTTGAAGCTTTGAGATTAAACAATGGCAGGAAATCATCTTTTAAAGTAAT 4964

QY 119 TCTCTTGGAGATGGTGGAGTTGGAAGAGTTTCTTATGAACAGATATGTAATAA 178  
 Db 4965 TCTCTTGGAGATGGTGGAGTTGGAAGAGTTTCTTATGAACAGATATGTAATAA 5024

QY 179 GTTTGATACCCAGCTCTTCCATACAATAGTGTGGAAATTTTAAATAAAGATTGGAAGT 238  
 Db 5025 GTTTGATACCCAGCTCTTCCATACAATAGTGTGGAAATTTTAAATAAAGATTGGAAGT 5084

QY 239 GGATGGACATTTTGTACCATGAGATTGGGACAGCGAGGTCAGGAGGATTCGGAAG 298  
 Db 5085 GGATGGACATTTTGTACCATGAGATTGGGACAGCGAGGTCAGGAGGATTCGGAAG 5144

QY 299 CCTGAGGACACCAATTTTACAGAGGTTCTGACTGCTGCTCTTACTTTTAGTGTGATGA 358  
 Db 5145 CCTGAGGACACCAATTTTACAGAGGTTCTGACTGCTGCTCTTACTTTTAGTGTGATGA 5204

QY 359 TTCAAAAGCTTCCAGAACTTAAGTAAGTGAAGAAAGATTTCATATATATTCAGATGT 418  
 Db 5205 TTCAAAAGCTTCCAGAACTTAAGTAAGTGAAGAAAGATTTCATATATATTCAGATGT 5264

QY 419 GAAAGAGCTTCCAGAGCTTCTTTTGTGATTTCTGGGTAAACAAGATTGACATAAGCGAAG 478  
 Db 5265 GAAAGAGCTTCCAGAGCTTCTTTTGTGATTTCTGGGTAAACAAGATTGACATAAGCGAAG 5324

QY 479 GCAGGTGTTACAGAAAGCCCAAGCTTGGTGAGGAGCAACGGGCACTATCTTTATTT 538  
 Db 5325 GCAGGTGTTACAGAAAGCCCAAGCTTGGTGAGGAGCAACGGGCACTATCTTTATTT 5384

QY 539 TGAACAAGTGCAAAAGATGCCCAATGTGGCAGAGCTTTTGAGGAGCGGTTCCGAG 598  
 Db 5385 TGAACAAGTGCAAAAGATGCCCAATGTGGCAGAGCTTTTGAGGAGCGGTTCCGAG 5444

QY 599 AGTTCTTGTACCGAGATAGGTGAGATCAATTTGATTCAGACAGACACAGTCAATCTTCA 658  
 Db 5445 AGTTCTTGTACCGAGATAGGTGAGATCAATTTGATTCAGACAGACACAGTCAATCTTCA 5504

QY 659 CCGAAGCCCAAGCCTAGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 718  
 Db 5505 CCGAAGCCCAAGCCTAGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5564

QY 719 CAACTCACATATACAAAATCAACATGGGATGGAGAGAAATTAGCGTTTTCGAG 778  
 Db 5565 CAACTCACATATACAAAATCAACATGGGATGGAGAGAAATTAGCGTTTTCGAG 5624

QY 779 AGTGTATCATCTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 838  
 Db 5625 AGTGTATCATCTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5684

QY 839 GGGACACATCCCTCTTGGAGGATATATTTACTCAATATGTCACCTTACATTTATATA 898  
 Db 5685 GGGACACATCCCTCTTGGAGGATATATTTACTCAATATGTCACCTTACATTTATATA 5744

QY 899 TTGTACAGTGTCTAATAACGTTTCTTTAATTAATAATATGTAAGTTCAGAGCTAATAA 958  
 Db 5745 TTGTACAGTGTCTAATAACGTTTCTTTAATTAATAATATGTAAGTTCAGAGCTAATAA 5804

QY 959 ATGAATACCCAGAGCTTTAATATATAAATAAATAAATAAATAAATAAATAAATAA 1018  
 Db 5805 ATGAATACCCAGAGCTTTAATATATAAATAAATAAATAAATAAATAAATAAATAA 5864

QY 1019 TACTTGGATTTTCTTGGGAAATGGAGAACTACTTTTATATGATGATGTTTATGTC 1078  
 Db 5865 TACTTGGATTTTCTTGGGAAATGGAGAACTACTTTTATATGATGATGTTTATGTC 5924

QY 1079 AATTAGCATGTATCTTGGTTCAGGAAATACCTTTCCCTAAAGCAATAATGTAAGTAT 1138

Db 5925 AATTAGCATGTATCTTGGTTCAGGAAATACCTTTCCCTAAAGCAATAATGTTAGATATT 5984

QY 1139 AAAGATTAAATCTAATGTA 1158  
 Db 5985 AAAGATTAAATCTAATGTA 6004

RESULT 10  
 ABA50249  
 ID ABA50249 standard; DNA; 632 BP.  
 XX  
 AC ABA50249;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE Human breast cell single exon nucleic acid probe #8944.  
 XX  
 KW Human; microarray; single exon probe; gene expression; breast;  
 KW disease; cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157271-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00662.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-496933/54.  
 XX  
 PT New spatially-addressable set of single exon nucleic acid probes,  
 PT useful for measuring gene expression in sample derived from human  
 PT breast, comprises number of single exon nucleic acid probes -  
 XX  
 PS Claim 4; SEQ ID NO 8944; 327bp + sequence listing; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and Br 474 cells. The method involves contacting  
 CC the probes with a collection of detectably labelled nucleic acids  
 CC derived from mRNA of human breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins. They are useful for gene discovery, and for  
 CC determining predisposition and/or prognosing breast disease. Gene  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression, with far less bias  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequence. The  
 CC present sequence is a single exon nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 632 BP; 190 A; 115 C; 153 G; 174 T; 0 other;

Query March 53.8%; Score 632; DB 22; Length 632;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-127;  
 Matches 632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 61 GGTCTTGAAGCTTTTGAATTAACATGCGAGGAAATCATCTTTTAAAGTAATTC 120  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1 GGTCTTGAAGCTTTTGAATTAACATGCGAGGAAATCATCTTTTAAAGTAATTC 60  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 121 TCCTTGGAGATGGTGGAGTTGGGAAGTTCCTTATGAAACAGATATGTAATAAGT 180  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 61 TCCTTGGAGATGGTGGAGTTGGGAAGTTCCTTATGAAACAGATATGTAATAAGT 120  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 181 TTGATACCCAGCTCTTCCATACATATAGGTGTGGATTTTAAATTAAGTTCGAGTGG 240  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 121 TTGATACCCAGCTCTTCCATACATATAGGTGTGGATTTTAAATTAAGTTCGAGTGG 180  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 241 ATGGACATTTTGTACCATGCGATTTGGGACACGCGAGTTCAGAGCGATTCGGAAGCC 300  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 181 ATGGACATTTTGTACCATGCGATTTGGGACACGCGAGTTCAGAGCGATTCGGAAGCC 240  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 301 TGAGGACACATTTTACAGAGTTCTGACTGCTGCTTACTTTTAAATTAAGTTCGAGTGG 360  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 241 TGAGGACACATTTTACAGAGTTCTGACTGCTGCTTACTTTTAAATTAAGTTCGAGTGG 300  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 361 CACAAAGCTTCCAGAGCTTAAAGTAACTGGAAGAAAGATTCATATATTATGAGATGTA 420  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 301 CACAAAGCTTCCAGAGCTTAAAGTAACTGGAAGAAAGATTCATATATTATGAGATGTA 360  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 421 AAGAGCTGAGAGCTTTCTTTTGTGATTTCTGGTAAACAGATTCACATTAAGTTCGAGTGG 480  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 361 AAGAGCTGAGAGCTTTCTTTTGTGATTTCTGGTAAACAGATTCACATTAAGTTCGAGTGG 420  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 481 AGGTGCTACAGAGAGCCAGCTTGGTGGACACGCGAGTTCAGAGCGATTCGGAAGCC 540  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 421 AGGTGCTACAGAGAGCCAGCTTGGTGGACACGCGAGTTCAGAGCGATTCGGAAGCC 480  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 541 AAACAAAGTGCAAAAGATGCCACAAATGTGGCAGCAGCTTTGAGGAGCGGTTGGAAGAG 600  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 481 AAACAAAGTGCAAAAGATGCCACAAATGTGGCAGCAGCTTTGAGGAGCGGTTGGAAGAG 540  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 601 TTCTTGCTACCGAGGTAGTGCAGATCTTTGATTCAGACACAGTCAATCTTCACC 660  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 541 TTCTTGCTACCGAGGTAGTGCAGATCTTTGATTCAGACACAGTCAATCTTCACC 600  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 661 GAAAGCCCAAGCTAGCTCATCTTGTGTTGA 692  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 601 GAAAGCCCAAGCTAGCTCATCTTGTGTTGA 632  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 11  
AAK16569  
ID AAK16569 standard; DNA; 632 BP.  
AC AAK16569;  
XX AAK16569;  
XX AAK16569;  
DT 05-NOV-2001 (first entry)  
XX Human brain expressed single exon probe SEQ ID NO: 16560.  
DE Human; brain expressed exon; gene expression analysis; probe;  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX Epilepsy; Cancer; ss.  
XX Homo sapiens.  
OS Homo sapiens.  
XX WO200157275-A2.  
XX PD 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00667.  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX MPI; 2001-483446/52.  
XX Single exon nucleic acid probes for analyzing gene expression in human brains -  
XX Example 4; SEQ ID NO: 16560; 650pp + Sequence Listing; English.  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX Sequence 632 BP; 190 A; 115 C; 153 G; 174 T; 0 other;  
SQ Query Match 53.8%; Score 632; DB 22; Length 632;  
Best Local Similarity 100.0%; Pred. No. 6.4e-127;  
Matches 632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 61 GGTCTTGAAGCTTTTGAATTAACATGCGAGGAAATCATCTTTTAAAGTAATTC 120  
Db 1 GGTCTTGAAGCTTTTGAATTAACATGCGAGGAAATCATCTTTTAAAGTAATTC 60  
QY 121 TCCTTGGAGATGGTGGAGTTGGGAAGTTCCTTATGAAACAGATATGTAATAAGT 180  
Db 61 TCCTTGGAGATGGTGGAGTTGGGAAGTTCCTTATGAAACAGATATGTAATAAGT 120  
QY 181 TTGATACCCAGCTCTTCCATACATATAGGTGTGGATTTTAAATTAAGTTCGAGTGG 240  
Db 121 TTGATACCCAGCTCTTCCATACATATAGGTGTGGATTTTAAATTAAGTTCGAGTGG 180  
QY 241 ATGGACATTTTGTACCATGCGATTTGGGACACGCGAGTTCAGAGCGATTCGGAAGCC 300  
Db 181 ATGGACATTTTGTACCATGCGATTTGGGACACGCGAGTTCAGAGCGATTCGGAAGCC 240  
QY 301 TGAGGACACATTTTACAGAGTTCTGACTGCTGCTTACTTTTAAATTAAGTTCGAGTGG 360  
Db 241 TGAGGACACATTTTACAGAGTTCTGACTGCTGCTTACTTTTAAATTAAGTTCGAGTGG 300  
QY 361 CACAAAGCTTCCAGAGCTTAAAGTAACTGGAAGAAAGATTCATATATTATGAGATGTA 420  
Db 301 CACAAAGCTTCCAGAGCTTAAAGTAACTGGAAGAAAGATTCATATATTATGAGATGTA 360  
QY 421 AAGAGCTGAGAGCTTTCTTTTGTGATTTCTGGTAAACAGATTCACATTAAGTTCGAGTGG 480  
Db 361 AAGAGCTGAGAGCTTTCTTTTGTGATTTCTGGTAAACAGATTCACATTAAGTTCGAGTGG 420  
QY 481 AGGTGCTACAGAGAGCCAGCTTGGTGGACACGCGAGTTCAGAGCGATTCGGAAGCC 540  
Db 421 AGGTGCTACAGAGAGCCAGCTTGGTGGACACGCGAGTTCAGAGCGATTCGGAAGCC 480  
QY 541 AAACAAAGTGCAAAAGATGCCACAAATGTGGCAGCAGCTTTGAGGAGCGGTTGGAAGAG 600  
Db 481 AAACAAAGTGCAAAAGATGCCACAAATGTGGCAGCAGCTTTGAGGAGCGGTTGGAAGAG 540  
QY 601 TTCTTGCTACCGAGGTAGTGCAGATCTTTGATTCAGACACAGTCAATCTTCACC 660  
Db 541 TTCTTGCTACCGAGGTAGTGCAGATCTTTGATTCAGACACAGTCAATCTTCACC 600  
QY 661 GAAAGCCCAAGCTAGCTCATCTTGTGTTGA 692  
Db 601 GAAAGCCCAAGCTAGCTCATCTTGTGTTGA 632

AAI23103  
ID AAI23103 standard; DNA; 632 BP.  
XX  
AC AAI23103;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Probe #13036 for gene expression analysis in human cervical cell sample.  
XX  
DE Probe; human; microarray; gene expression; cervical epithelial cell;  
XX  
KW cervical cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157278-A2.  
XX  
PD 09-AUG-2001.  
XX  
PP 30-JAN-2001; 2001WO-US00670.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SEP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488901/53.  
XX  
DR Human genome-derived single exon nucleic acid probes useful for  
XX  
PT analyzing gene expression in human cervical epithelial cells -  
XX  
PS Claim 25; SEQ ID No 13036; 487pp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
CC (SENP). The present sequence is one such probe. The SENPs are derived  
CC from human Hela cells. The SENPs can be used to produce a single exon  
CC microarray, which can be used for measuring human gene expression in a  
CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging  
CC of diseases of the cervix, notably cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 632 BP; 190 A; 115 C; 153 G; 174 T; 0 other;  
  
Query Match 53.8%; Score 632; DB 22; Length 632;  
Best Local Similarity 100.0%; Pred. No. 6.4e-127;  
Matches 632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 61 GGTCTTGAAGCTTTGAGATTAACTGCGAGGAATCATCACTTTTAAAGTAATTC 120  
DB 1 GGTCTTGAAGCTTTGAGATTAACTGCGAGGAATCATCACTTTTAAAGTAATTC 60  
  
QY 121 TCCTTGGAGATGGTGGAGTTGGGAAGATTCCTATTAACACAGATATGTAATAAAGT 180  
DB 61 TCCTTGGAGATGGTGGAGTTGGGAAGATTCCTATTAACACAGATATGTAATAAAGT 120  
  
QY 181 TTGATACCCAGCTCTCCATACAAATAGGTGGGATTTTAAATAAAGATTGGAGTGG 240  
DB 121 TTGATACCCAGCTCTCCATACAAATAGGTGGGATTTTAAATAAAGATTGGAGTGG 180  
  
QY 241 ATGGACATTTTGTACCATGAGATTGGACACGCGAGGTGAGAGGATTCCGAAGCC 300  
DB 181 ATGGACATTTTGTACCATGAGATTGGACACGCGAGGTGAGAGGATTCCGAAGCC 240  
  
QY 301 TGAGGACACCATTTACAGAGGTTCTGACTGCTGCTTACTTTTAGTCGATGATT 360

DB 241 TGAGGACACCATTTACAGAGGTTCTGACTGCTGCTTACTTTTAGTCGATGATT 300  
QY 361 CACAAAGCTTCCAGAACTTAAAGTAACCTGGAAGAAAGAAATTCATATATATGAGATGTA 420  
DB 301 CACAAAGCTTCCAGAACTTAAAGTAACCTGGAAGAAAGAAATTCATATATATGAGATGTA 360  
QY 421 AAGAGCTGAGAGCTTTCCTTTTGTGATCTGCGTAACTGAGATTGACATAAGCGAAGCC 480  
DB 361 AAGAGCTGAGAGCTTTCCTTTTGTGATCTGCGTAACTGAGATTGACATAAGCGAAGCC 420  
QY 481 AGGTGCTTACAGAAAGCCCAAGCTTGTGCGAGGACAAACGCGACTATTCCTTTATTTG 540  
DB 421 AGGTGCTTACAGAAAGCCCAAGCTTGTGCGAGGACAAACGCGACTATTCCTTTATTTG 480  
QY 541 AATCAAGTGCAAAAGATGCCACAAATGTGCGACGACCTTTGAGGAAGCGGTTCCGAAG 600  
DB 481 AATCAAGTGCAAAAGATGCCACAAATGTGCGACGACCTTTGAGGAAGCGGTTCCGAAG 540  
QY 601 TTCTTGTCTACCGAGGATAGTTCAGATCATTTGATTTCAGACAGACACAGTCAATCTTCACC 660  
DB 541 TTCTTGTCTACCGAGGATAGTTCAGATCATTTGATTTCAGACAGACACAGTCAATCTTCACC 600  
QY 661 GAAAGCCCAAGCTTAGCTCATCTTCTGCTGTTGA 692  
DB 601 GAAAGCCCAAGCTTAGCTCATCTTCTGCTGTTGA 632  
  
RESULT 13  
AAI08751  
ID AAI08751 standard; DNA; 632 BP.  
XX  
AC AAI08751;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Probe #8742 used to measure gene expression in human breast sample.  
XX  
XX Probe; human; breast disease; breast cancer; development disorder; ss;  
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
OS Homo sapiens.  
XX  
PN WO200157270-A2.  
XX  
PD 09-AUG-2001.  
XX  
PP 29-JAN-2001; 2001WO-US00661.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SEP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-476286/51.  
XX  
DR Novel single exon nucleic acid probe used to measuring gene expression  
XX in a human breast -  
XX  
PS Claim 25; SEQ ID No 8742; 322pp; English.  
XX  
XX The present invention relates to novel single exon nucleic acid probes.  
CC The present sequence is one such probe. The probes are useful for  
CC measuring human gene expression in a human breast sample, where the probe  
CC hybridises at high stringency to a nucleic acid expressed in the human  
CC breast. The probes are useful for predicting, diagnosing, grading,  
CC

CC staging, monitoring and prognosing diseases of the human breast.  
 CC particularly those diseases with polygenic aetiology. The diseases  
 CC include: breast cancer, disorders of development, inflammatory diseases  
 CC of the breast, fibrocystic changes, proliferative breast disease and  
 CC non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 SQ Sequence 632 BP; 190 A; 115 C; 153 G; 174 T; 0 other;

Query Match 53.8%; Score 632; DB 22; Length 632;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-127;  
 Matches 632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 GGTTCTTGAAGCTTTTGAATTAACATGCGAGGAAATCATCACTTTTAAAGTAATTC 120  
 Db 1 GGTTCTTGAAGCTTTTGAATTAACATGCGAGGAAATCATCACTTTTAAAGTAATTC 60

QY 121 TCCTTGGAGATGGTGGAGTTGGAGAGTTCCTTATGAAACAGATATGTAATTAAGT 180  
 Db 1 TCCTTGGAGATGGTGGAGTTGGAGAGTTCCTTATGAAACAGATATGTAATTAAGT 120

QY 181 TTGATACCCAGCTTCCATACATAGTGTGGAATTTTAAATTAAGTTGGAGTGG 240  
 Db 1 TTGATACCCAGCTTCCATACATAGTGTGGAATTTTAAATTAAGTTGGAGTGG 180

QY 241 ATGGACATTTTGTACCATTCAGATTTGGGACACGCGAGTTCAGGAGGATTTCCGAAGCC 300  
 Db 1 ATGGACATTTTGTACCATTCAGATTTGGGACACGCGAGTTCAGGAGGATTTCCGAAGCC 240

QY 301 TGAGGACACATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTGTGTCGATGTT 360  
 Db 1 TGAGGACACATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTGTGTCGATGTT 300

QY 361 CACAAAGCTTCCAGAACTTAAAGTAACTGGAACAAAGAAATTCATATATTATGAGATGTA 420  
 Db 1 CACAAAGCTTCCAGAACTTAAAGTAACTGGAACAAAGAAATTCATATATTATGAGATGTA 360

QY 421 AAGAGCTGAGAGCTTTCTTTTGTGATTTCTGGTAAACAGATTGACATAGCGAAGCGC 480  
 Db 1 AAGAGCTGAGAGCTTTCTTTTGTGATTTCTGGTAAACAGATTGACATAGCGAAGCGC 420

QY 481 AGTGCTGTACAGAAAGCCCAAGCTTGTGTGAGGACCAAGCGGACTATCTTATTTTG 540  
 Db 1 AGTGCTGTACAGAAAGCCCAAGCTTGTGTGAGGACCAAGCGGACTATCTTATTTTG 480

QY 541 AAACAGTGCAGAAAGATGCCCAAAATGTGGCAGAGCCTTTGAGGAAGCGGTTGAGAG 600  
 Db 1 AAACAGTGCAGAAAGATGCCCAAAATGTGGCAGAGCCTTTGAGGAAGCGGTTGAGAG 540

QY 601 TTCTTGCTACCGAGATAGGTGAGATCATTGATTCAGACAGACAGATCACTTTCACC 660  
 Db 1 TTCTTGCTACCGAGATAGGTGAGATCATTGATTCAGACAGACAGATCACTTTCACC 600

QY 661 GAAAGCCCAAGCTAGCTCATCTTGTGTTGA 692  
 Db 1 GAAAGCCCAAGCTAGCTCATCTTGTGTTGA 632

RESULT 14  
 ABS16384  
 ID ABS16384 standard; DNA; 632 BP.

XX ABS16384;  
 AC ABS16384;  
 DT 19-AUG-2002 (first entry)  
 XX Human genome-derived single exon probe ORF from lung SEQ ID No 16375.  
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease; open reading frame; ORF.  
 OS Homo sapiens.  
 XX WO2001186003-A2.  
 XX 15-NOV-2001.  
 XX 30-JAN-2001; 2001WO-US00665.  
 XX 04-FEB-2000; 2000US-180312P.  
 XX 26-MAY-2000; 2000US-207456P.  
 XX 30-JUN-2000; 2000US-0608408.  
 XX 03-AUG-2000; 2000US-0623366.  
 XX 21-SEP-2000; 2000US-234687P.  
 XX 27-SEP-2000; 2000US-236359P.  
 XX 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DX, Chen W, Rank DR;  
 PI WPI; 2002-114183/15.  
 DR Spatially-addressable set of single exon nucleic acid probes, used to  
 XX measure gene expression in human lung samples -  
 PT Claim 4; SEQ ID No 16375; 634pp; English.

The invention relates to a spatially-addressable set of single exon  
 nucleic acid probes for measuring gene expression in a sample derived  
 from human lung comprising single exon nucleic acid probes having one of  
 1614 nucleic acid sequences mentioned in the specification, or their  
 complements or the 12387 open reading frames derived from the 12614  
 probes. Also included are a microarray comprising the novel set of  
 probes. The novel set of probes which hybridise at high stringency to a  
 nucleic acid expressed in the human lung; measuring gene expression in a  
 sample derived from human lung, comprising (a) contacting the array with  
 a collection of detectably labeled nucleic acids derived from human lung  
 mRNA, and (b) measuring the label detectably bound to each probe of  
 the array; identifying exons in a eukaryotic genome, comprising  
 (a) algorithmically predicting at least one exon from genomic sequences  
 of the eukaryote; and (b) detecting specific hybridisation of detectably  
 labeled nucleic acids from eukaryote lung mRNA, to a single exon probe.  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberosus sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a single exon  
 CC probe open reading frame of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at

```

CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 632 BP; 190 A; 115 C; 153 G; 174 T; 0 other;

Query Match      53.8%; Score 632; DB 24; Length 632;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 61 GGTCTTGAAGCTTTTGGAGATTAACAAATGGCAGGAAATCATCATCTTTTAAAGTAATTC 120
Db 1 GGTCTTGAAGCTTTTGGAGATTAACAAATGGCAGGAAATCATCATCTTTTAAAGTAATTC 60
Qy 121 TCCTTGGAGATGGTGGAGTGGGAGAGTTCACCTTATGAACAGATATGAACCTAATAGT 180
Db 61 TCCTTGGAGATGGTGGAGTGGGAGAGTTCACCTTATGAACAGATATGAACCTAATAGT 120
Qy 181 TTGATACCCAGCTCTCCATACATAGCTGGAATTTTAAATAAGATTTGGAGTGG 240
Db 121 TTGATACCCAGCTCTCCATACATAGCTGGAATTTTAAATAAGATTTGGAGTGG 180
Qy 241 ATGGACATTTTGTATACCATCAGATTTGGACACGGCAGGTCCAGGCGCATTCGGAAGCC 300
Db 181 ATGGACATTTTGTATACCATCAGATTTGGACACGGCAGGTCCAGGCGCATTCGGAAGCC 240
Qy 301 TGAGGACACCATTTTACAGAGTCTGACTGCTGCCCTTACTTTTAGTTCGATGAT 360
Db 241 TGAGGACACCATTTTACAGAGTCTGACTGCTGCCCTTACTTTTAGTTCGATGAT 300
Qy 361 CACAAGCTTCCAGAACTTAAAGTAAGTAACTGGAAGAAAGAAATTCATATATTGCGAGTGCA 420
Db 301 CACAAGCTTCCAGAACTTAAAGTAAGTAACTGGAAGAAAGAAATTCATATATTGCGAGTGCA 360
Qy 421 AAGAGCTGAGAGCTTCTTTTGTGATCTGCGTAAAGATGACATGAAGTGAATGAAGCGGAC 480
Db 361 AAGAGCTGAGAGCTTCTTTTGTGATCTGCGTAAAGATGACATGAAGTGAATGAAGCGGAC 420
Qy 481 AGGTGCTACAGAAAGCCCAAGCTTGGTGGAGGACACAGCGGACTATCTTATTGTTG 540
Db 421 AGGTGCTACAGAAAGCCCAAGCTTGGTGGAGGACACAGCGGACTATCTTATTGTTG 480
Qy 541 AAACAAGTGCAAAAGATGCGCAAAATGTGGCAGCAGCCTTTGAGGAAGCGGTCGAAGAG 600
Db 481 AAACAAGTGCAAAAGATGCGCAAAATGTGGCAGCAGCCTTTGAGGAAGCGGTCGAAGAG 540
Qy 601 TTCTTGTACCGAGGATAGTACAGATCTTGTGATTCAGACAGACACAGTCATCTTCACC 560
Db 541 TTCTTGTACCGAGGATAGTACAGATCTTGTGATTCAGACAGACACAGTCATCTTCACC 500
Qy 661 GAAAGCCCAAGCCTAGCTCATCTTGTGTTGA 692
Db 601 GAAAGCCCAAGCCTAGCTCATCTTGTGTTGA 632

RESULT 15
ACA56484
ID ACA56484 standard; cdna; 606 BP.
XX
AC ACA56484;
XX
AC ACA56484;
XX
DT 06-JUN-2003 (first entry)
XX
DE Human signalling pathway polynucleotide probe SEQ ID NO 1082.
XX
KW Human; probe; ss; array element; Parkinson's disease;
XX
KW signalling pathway population; cancer; adenocarcinoma; leukaemia;
XX
KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
XX
OS Homo sapiens.
XX
PN US6500938-B1.
XX
PD 31-DEC-2002.
XX

```

```

PF 30-JAN-1998; 98US-0016434.
XX
PR 30-JAN-1998; 98US-0016434.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Au-Young J, Seilhamer JJ;
XX
DR WPI; 2003-352189/33.
XX
PT Combination of polynucleotide probes, useful as array elements in a
PT microarray for monitoring the expression of a number of target
PT polynucleotides
XX
PS Claim 1; SEQ ID NO 1082; 65pp; English.
XX
CC The invention relates to a combination which, comprises a number of
CC polynucleotide probes comprising a sequence selected from one of the 1490
CC sequences mentioned in the specification. The combination is useful as an
CC array element in a microarray for monitoring the expression of a number
CC of target polynucleotides. The microarray is particularly useful in the
CC diagnosis and treatment of cancer and immunopathology and neuropathology.
CC The microarray is useful in diagnostics and treatment regimens, drug
CC discovery and development, toxicological and carcinogenicity studies,
CC forensics and pharmacogenomics. The microarray is also useful for
CC monitoring progression of diseases and for developing sophisticated
CC profiles for the effects of currently available therapeutic drugs. The
CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs
CC and genomic fragments and in research and diagnostic applications. The
CC array can detect changes in expression in a large number of genes coding
CC for different signaling pathway populations which can be used to diagnose
CC various diseases including cancer e.g. adenocarcinoma and leukaemia,
CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
CC and Parkinson's disease. The present sequence represents a polynucleotide
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=06500938B1.
XX
SQ Sequence 606 BP; 182 A; 112 C; 147 G; 165 T; 0 other;

Query Match      51.4%; Score 604.4; DB 25; Length 606;
Best Local Similarity 99.8%; Pred. No. 5.7e-121;
Matches 605; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 87 ATGGCAGGAAATCATCACTTTTAAAGTAATTCCTCTGGAGATGGTGGAGTGGGAAG 146
Db 1 ATGGCAGGAAATCTTCACITTTTAAAGTAATTCCTCTGGAGATGGTGGAGTGGGAAG 60
Qy 147 AGTTCACATTATGAACAGATATGTAACCTAATTAAGTTGATACCCAGCTCTCCATACATA 206
Db 61 AGTTCACATTATGAACAGATATGTAACCTAATTAAGTTGATACCCAGCTCTCCATACATA 120
Qy 207 GGTGTGGAATTTTAAATAAAGATTTGGAAGTGGATGACATTTTGTATCCATCAGATT 266
Db 121 GGTGTGGAATTTTAAATAAAGATTTGGAAGTGGATGACATTTTGTATCCATCAGATT 180
Qy 267 TGGGACACGGCAGGTCCAGGCGGATTCGGAAGCCTGAGGACACCATTTTACAGAGTTCT 326
Db 181 TGGGACACGGCAGGTCCAGGCGGATTCGGAAGCCTGAGGACACCATTTTACAGAGTTCT 240
Qy 327 GACTGCTGCCCTGCTTACTTTTAGTGTGATGATTCACAAAGCTTCCAGAACTTAAGTAAC 386
Db 241 GACTGCTGCCCTGCTTACTTTTAGTGTGATGATTCACAAAGCTTCCAGAACTTAAGTAAC 300
Qy 397 TGGGAAGAAAGATTCATATATATATGACAGATGTGGAAGCGCTGAGAGCTTTCCTTTGTG 446
Db 301 TGGGAAGAAAGATTCATATATATATGACAGATGTGGAAGCGCTGAGAGCTTTCCTTTGTG 360
Qy 447 ATCTTGCTTACAGATTTGACATAAGCGGCGGAGGTGTCTACAGAGAGGCCCAAGCT 506
Db 361 ATCTTGCTTACAGATTTGACATAAGCGGCGGAGGTGTCTACAGAGAGGCCCAAGCT 420

```

QY	507	TGCTGAGGACAAACGGCGACTATCCTTATTTTGAACAAGTGCAAAAGATGCCACAAT	566
Db	421	TGCTGAGGACAAACGGCGACTATCCTTATTTTGAACAAGTGCAAAAGATGCCACAAT	480
QY	567	GTGGCAGCAGCCTTTGAGGAAGCGTTGGAAGAGTTCTTGCTACCGAGGATAGGTCAGAT	626
Db	481	GTGGCAGCAGCCTTTGAGGAAGCGTTGGAAGAGTTCTTGCTACCGAGGATAGGTCAGAT	540
QY	627	CATTGATTCAGACACACAGTCATCTTCACCGAAGCGCCAGGCTAGCTCATCTTGC	686
Db	541	CATTGATTCAGACACACAGTCATCTTCACCGAAGCGCCAGGCTAGCTCATCTTGC	600
QY	687	TGTTGA	692
Db	601	TGTTGA	606

Search completed: February 10, 2004, 02:39:12  
Job time : 396 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2004, 01:51:16 ; Search time 92 Seconds  
(without alignments)  
5637.230 Million cell updates/sec

Title: US-09-988-974-6

Perfect score: 1175

Sequence: 1 CACTGTGATGAACACTTTT.....GTAAAAA.....1175

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 113956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgm2\_6/ptodata/2/ina/5A COMB.seq.\*
- 2: /cgm2\_6/ptodata/2/ina/5B COMB.seq.\*
- 3: /cgm2\_6/ptodata/2/ina/6A COMB.seq.\*
- 4: /cgm2\_6/ptodata/2/ina/6B COMB.seq.\*
- 5: /cgm2\_6/ptodata/2/ina/PTUS COMB.seq.\*
- 6: /cgm2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1175	100.0	1175	2	US-08-773-423-6
2	604.4	51.4	606	4	US-09-016-434-1082
3	439.6	37.4	487	3	US-09-030-607-206
4	439.6	37.4	487	4	US-09-439-313-206
5	439.6	37.4	487	4	US-09-352-618A-206
6	439.6	37.4	487	4	US-09-232-149A-206
7	306.2	26.1	473	3	US-09-020-956-106
8	306.2	26.1	473	3	US-09-030-607-106
9	306.2	26.1	473	4	US-09-439-313-106
10	306.2	26.1	473	4	US-09-352-616A-106
11	306.2	26.1	473	4	US-09-232-149A-106
12	181.8	15.5	624	4	US-09-016-434-1083
13	98.8	8.4	970	3	US-08-888-077A-28
14	81.2	6.9	723	4	US-09-016-434-1422
15	69.6	5.9	944	4	US-09-016-434-1060
16	65.8	5.6	1069	4	US-09-620-312D-646
17	63.8	5.4	1533	4	US-09-075-454-11
18	63.6	5.4	1172	4	US-09-075-454-8
19	63.6	5.4	1255	2	US-08-766-551-6
20	62.8	5.3	986	4	US-09-620-312D-647
21	62.8	5.3	925	2	US-08-916-901-4
22	62.8	5.3	925	4	US-09-154-602-4
23	60.2	5.1	286	4	US-09-313-294A-6296
24	60.2	5.1	651	4	US-09-016-434-1255
25	59.8	5.1	850	4	US-09-669-751-150
26	59.4	5.1	875	4	US-09-075-454-10
27	59.4	5.1	1106	4	US-09-620-312D-959

28	59.4	5.1	2612	4	US-09-484-970B-142	Sequence 142, Appl
29	58.6	5.0	730	4	US-09-300-958A-16	Sequence 16, Appl
30	58	4.9	639	4	US-09-399-313-66	Sequence 66, Appl
31	55.4	4.7	1631	4	US-09-620-312D-587	Sequence 587, Appl
32	55	4.7	472	4	US-09-325-932A-30	Sequence 30, Appl
33	53.8	4.6	833	4	US-09-620-312D-426	Sequence 426, Appl
34	53.6	4.6	803	4	US-09-075-454-13	Sequence 13, Appl
35	53.2	4.5	1340	2	US-08-824-873-2	Sequence 2, Appl
36	53.2	4.5	1340	3	US-09-198-184-2	Sequence 2, Appl
37	53.2	4.5	1407	4	US-09-433-314-1	Sequence 1, Appl
38	52.8	4.5	847	2	US-08-773-423-4	Sequence 4, Appl
39	51.2	4.4	1102	4	US-09-620-312D-828	Sequence 828, Appl
40	51.2	4.4	1127	4	US-09-391-741A-1	Sequence 1, Appl
41	51.2	4.4	1127	4	US-09-391-741A-25	Sequence 25, Appl
42	51	4.3	241	4	US-09-016-434-813	Sequence 813, Appl
43	51	4.3	465	4	US-09-325-932A-34	Sequence 34, Appl
44	50	4.3	719	2	US-08-766-551-2	Sequence 2, Appl
45	50	4.3	719	4	US-09-016-434-76	Sequence 76, Appl

## ALIGNMENTS

RESULT 1  
US-08-773-423-6  
; Sequence 6, Application US/08773423  
; Patent No. 5869291  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL RAB PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/773,423  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0183 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1175 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Consensus  
; CLONE: Consensus  
US-08-773-423-6  
Query Match 100.0%; Score 1175; DB 2; Length 1175;  
Best Local Similarity 100.0%; Pred. No. 6e-286;

Matches 1175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACTGTGATGAACACTTTTCCCGTGTGCTTTGAGTGCATCTTCTCAACCAACCTTAGGAG 60

Db 1 CACTGTGATGAACACTTTTCCCGTGTGCTTTGAGTGCATCTTCTCAACCAACCTTAGGAG 60

QY 61 GGTTCCTGAGCTTTTGAATTAACAATGGCAGGAATATCATCTTTTAAAGTAAATTC 120

Db 61 GGTTCCTGAGCTTTTGAATTAACAATGGCAGGAATATCATCTTTTAAAGTAAATTC 120

QY 121 TCCTTGGAGATGGTGGAGTTGGGAAGATTCACATATGAACAGATATGTAACTAATAGT 180

Db 121 TCCTTGGAGATGGTGGAGTTGGGAAGATTCACATATGAACAGATATGTAACTAATAGT 180

QY 181 TTGATACCCAGCTTCTCCATACATAGGTGTGGATTTTAAATAAGATTTGGAAGTGG 240

Db 181 TTGATACCCAGCTTCTCCATACATAGGTGTGGATTTTAAATAAGATTTGGAAGTGG 240

QY 241 ATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTTCAGGAGCGATTCCGAAGCC 300

Db 241 ATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTTCAGGAGCGATTCCGAAGCC 300

QY 301 TGAGACACCATTTTACAGAGTCTGACTGCTGCTTACTTTTAGTGTGATGATT 360

Db 301 TGAGACACCATTTTACAGAGTCTGACTGCTGCTTACTTTTAGTGTGATGATT 360

QY 361 CACAAAGCTTCAGAACTTAAGTAACTGGAAGAAATTCATATATTCAGATGTA 420

Db 361 CACAAAGCTTCAGAACTTAAGTAACTGGAAGAAATTCATATATTCAGATGTA 420

QY 421 AAGAGCTGAGAGCTTCTCTTTTGTGATCTGGGTAAACAAGATTGACATAGCGAAGCGC 480

Db 421 AAGAGCTGAGAGCTTCTCTTTTGTGATCTGGGTAAACAAGATTGACATAGCGAAGCGC 480

QY 481 AGGTGCTTACAGAGAGCCCAAGCTTGTGTCAGGACCAAGCGGACTATCTTTATTG 540

Db 481 AGGTGCTTACAGAGAGCCCAAGCTTGTGTCAGGACCAAGCGGACTATCTTTATTG 540

QY 541 AAACAAGTGCAAAAGATGCCAAATGTGGCAGCAGCTTTGAGGAAGCGGTTCGAAGAG 600

Db 541 AAACAAGTGCAAAAGATGCCAAATGTGGCAGCAGCTTTGAGGAAGCGGTTCGAAGAG 600

QY 601 TTCTTGTACCGAGATAGGTGAGATCATTTGATTCACAGACACAGTCAATCTTCAAC 660

Db 601 TTCTTGTACCGAGATAGGTGAGATCATTTGATTCACAGACACAGTCAATCTTCAAC 660

QY 661 GAAAGCCCAAGCTAGCTCATCTTCTGCTGATTTAGTGTGATGATTTCAACCA 720

Db 661 GAAAGCCCAAGCTAGCTCATCTTCTGCTGATTTAGTGTGATGATTTCAACCA 720

QY 721 ACTCACATATACACAAATCAACATGGGATGGAGAGAGATTTAGCTTTCAGCAG 780

Db 721 ACTCACATATACACAAATCAACATGGGATGGAGAGAGATTTAGCTTTCAGCAG 780

QY 781 TGTATCATCTACTAATAAATAAATAATGCTTCTGCTTCAATAGTTGGTGGGAGAGG 840

Db 781 TGTATCATCTACTAATAAATAAATAATGCTTCTGCTTCAATAGTTGGTGGGAGAGG 840

QY 841 GACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCAGCTTACATTTAAT 900

Db 841 GACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCAGCTTACATTTAAT 900

QY 901 GTAACAGTTGTCTAATAAGCTTTCTTTTAAATATGTAAGTTGACAGCTAATAAT 960

Db 901 GTAACAGTTGTCTAATAAGCTTTCTTTTAAATATGTAAGTTGACAGCTAATAAT 960

QY 961 GAAATGACCAAGACTTTAATAATAAATAAATAAGAACTTGACTATTCTAGAGTTATA 1020

Db 961 GAAATGACCAAGACTTTAATAATAAATAAATAAGAACTTGACTATTCTAGAGTTATA 1020

QY 1021 CTTGGATTTTCTTGGGAAATGAGACACTTTTATATGCTATGCTTTTATGCA 1080

Db 1021 CTTGGATTTTCTTGGGAAATGAGACACTTTTATATGCTATGCTTTTATGCA 1080

QY 1081 TTAGCATTCATCTTCTGTTTCTAGGAAATACCTTTCTTAAGCAATAATGTTAGATATAA 1140

Db 1081 TTAGCATTCATCTTCTGTTTCTAGGAAATACCTTTCTTAAGCAATAATGTTAGATATAA 1140

QY 1141 AGATTAAATCTAATGTTAAATGTTAAATGTTAAATGTTAAATGTTAAATGTTAAAT 1175

Db 1141 AGATTAAATCTAATGTTAAATGTTAAATGTTAAATGTTAAATGTTAAATGTTAAAT 1175

RESULT 2

US-09-016-434-1082  
 ; Sequence 1082, Application US/09016434  
 ; Patent No. 6500938  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Janice Au-Young  
 ; APPLICANT: Jeffrey J. Seilhamer  
 ; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
 ; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
 ; NUMBER OF SEQUENCES: 1490  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
 ; STREET: 3174 PORTER DRIVE  
 ; CITY: PALO ALTO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/016,434  
 ; FILING DATE: HEREWITH  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Zeller, Karen J.  
 ; REGISTRATION NUMBER: 37,071  
 ; REFERENCE/DOCKET NUMBER: PA-0002 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (650) 855-0555  
 ; TELEFAX: (650) 845-4166  
 ; INFORMATION FOR SEQ ID NO: 1082:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 606 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: GENBANK  
 ; CLONE: g1174146  
 ; US-09-016-434-1082

Query Match 51.4%; Score 604.4; DB 4; Length 606;  
 Best Local Similarity 99.8%; Pred. No. 9.9e-143;  
 Matches 605; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 87 ATGGCAGGAAATCATCACTTTTAAAGTAATTCCTTGGAGATGGTGGAGTTGGGAAG 146

Db 1 ATGGCAGGAAATCTTCATCTTTTAAAGTAATTCCTTGGAGATGGTGGAGTTGGGAAG 60

QY 147 AGTTCACTTATGACAGATATGTAATAAGTTTGTATGATGATGATGATGATGATGAT 206

Db 61 AGTTCACTTATGACAGATATGTAATAAGTTTGTATGATGATGATGATGATGATGAT 120

QY 207 GGTGTGGAATTTTAAATAAGATTTTGAAGTGGATGAGATTTTGTATGATGATGAT 266

Db 121 GGTGTGGAATTTTAAATAAGATTTTGAAGTGGATGAGATTTTGTATGATGATGAT 180

QY 267 TGGGACCGCAGGTGAGGAGCGATCCGAGCGCTGAGGACACCACTTTTACAGAGTTCT 326  
Db 181 TGGGACACCGCAGGTGAGGAGCGATCCGAGCGCTGAGGACACCACTTTTACAGAGTTCT 240  
QY 327 GACTGCTGCTGCTTACTTTAGTGTGATGATTTCAAAAGCTTCCAGAACTTAAGTAAC 386  
Db 241 GACTGCTGCTGCTTACTTTAGTGTGATGATTTCAAAAGCTTCCAGAACTTAAGTAAC 300  
QY 387 TGAAGAAGAAATCATATATTATGAGATGTAAGAGCTGAGAGCTTCTCTTTTGTG 446  
Db 301 TGAAGAAGAAATTCATATATTATGAGATGTAAGAGCTGAGAGCTTCTCTTTTGTG 360  
QY 447 ATTCTGGGTAAACAAGATTGATTAAGCGACGCGAGTGTCTACGAAGAAGCCCAAGCT 506  
Db 361 ATTCTGGGTAAACAAGATTGATTAAGCGACGCGAGTGTCTACGAAGAAGCCCAAGCT 420  
QY 507 TGTGACGGGACAAAGCGGCTATCTTTATTTTGAACAAGTGCACAAAGATGCCCAAT 566  
Db 421 TGTGACGGGACAAAGCGGCTATCTTTATTTTGAACAAGTGCACAAAGATGCCCAAT 480  
QY 567 GTGCGACGCGCTTGGAGGAGCGTTTGAAGATTCTTGTACCGAGGATAGGTGAGT 626  
Db 481 GTGCGACGCGCTTGGAGGAGCGTTTGAAGATTCTTGTACCGAGGATAGGTGAGT 540  
QY 627 CATTTGATTCAGACAGACAGTCAATCTTTCACCGAAAGCCCAAGCTAGCTCATCTTGC 686  
Db 541 CATTTGATTCAGACAGACAGTCAATCTTTCACCGAAAGCCCAAGCTAGCTCATCTTGC 600  
QY 687 TGTGTA 692  
Db 601 TGTGTA 606

RESULT 3  
US-09-030-607-206/c  
; Sequence 206, Application US/09030607  
; Patent No. 6262245  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO  
; NUMBER OF SEQUENCES: 224  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/030,607  
; FILING DATE: 25-FEB-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.427C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 206:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 487 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna

US-09-030-607-206  
Query Match 37.4%; Score 439.6; DB 3; Length 487;  
Best Local Similarity 96.0%; Pred. No. 2.3e-101;  
Matches 453; Conservative 0; Mismatches 18; Indels 1; Gaps 1;  
QY 536 TTTTGAACAAGTGCACAAAGATGCCACAAATGTGGCAGCGCTTTCAGGAAGCGGTTCG 595  
Db 487 TTTTGAACAAGTGCACAAAGATGCCACAAATGTGGCAGCGCTTTCAGGAAGCGGTTCG 428  
QY 596 AAGAGTTCTTCTACCGAGGATAGGTGAGATCAATTTGATTGAGATGTTGATGATTC 655  
Db 427 AAGAGTTCTTCTACCGAGGATAGGTGAGATCAATTTGATTGAGATGTTGATGATTC 368  
QY 656 TCACCGAAAGCCCAAGCGCTAGC-TCATCTTCTGCTGCTGATTGTTAGATGTTGATTC 714  
Db 367 TCACCGAAAGCCCAAGCGCTAGCTTCATCTTCTGCTGCTGATTGTTAGATGTTGATTC 308  
QY 715 TAACCAACTCACACATATACACAAATCAACATGGGATGGAGAGAGAATTAGCGTTTG 774  
Db 307 TNACCAACTCACACATATACACATATACACATATACACATATACACATATACCGTTTG 248  
QY 775 CAGCAGTGTATCATCTACTACTATAAAATTAACCTAATGTTGCTGCTTCAATGTTGTTGG 834  
Db 247 CAGCAGTGTATNATCTACTATAAAATTAACCTAATGTTGCTGCTTCAATGTTGTTGG 188  
QY 835 AGAAGGACACATCCACTCTTGGAGGAATATATTTACTCAATATGCGACCTTACATTTA 894  
Db 187 AGAAGGACACATCCACTCTTGGAGGAATATATNATCTCAATATGCGACCTTACATTTA 128  
QY 895 TAAATTGTAACAGTTGTCTAATAACGTTTCTTTAATTAATATGTAAGTTGCAAGCTA 954  
Db 127 TAAATTGTAACAGTTGTCTAATAACGTTTCTTTAATTAATATGTAAGTTGCAAGCTA 68  
QY 955 ATAAATGAATGACCAAGCTTTAATTAATTAATAAATAAGAACTTGACTA 1006  
Db 67 ATAAATGAATGACCAAGCTTTAATTAATTAATAAATAAGAACTTGACTA 16

RESULT 4  
US-09-439-313-206/c  
; Sequence 206, Application US/09439313  
; Patent No. 6529505  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan Louise  
; APPLICANT: Jiang Yuqui  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Kalos, Michael  
; APPLICANT: Fanger, Gary  
; APPLICANT: Retter, Mark  
; APPLICANT: Solk, John  
; APPLICANT: Day, Craig  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
; FILE REFERENCE: 210121.427C9  
; CURRENT APPLICATION NUMBER: US/09/439,313  
; CURRENT FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 575  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 206  
; LENGTH: 487  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(487)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-439-313-206  
Query Match 37.4%; Score 439.6; DB 4; Length 487;



Best Local Similarity 96.0%; Pred. No. 2.3e-101; Indels 1; Gaps 1;  
Matches 453; Conservative 0; Mismatches 18;

QY 536 TTTTGAACAAGTGCACCAAGATGCCACAAATGTGGCAGAGCCTTTGAGGAAGCGGTTCG 595  
Db 487 TTTTGAACAAGTGCACCAAGATGCCACAAATGTGGCAGAGCCTTTGAGGAAGCGGTTCG 428

QY 596 AAGAGTCTTCTACCGAGGATAGTGCAGATCATTTGATTTAGATTGTTGATGCAATCT 655  
Db 427 AAGAGTCTTCTACCGAGGATAGTGCAGATCATTTGATTTAGATTGTTGATGCAATCT 368

QY 656 TCACCGAAAGCCCAAGCCTAGC-TCATCTTCTGCTGTGATTTAGATTGTTGATGCAATCT 714  
Db 367 TCACCGAAAGCCCAAGCCTAGCCTTCTGCTGTGATTTAGATTGTTGATGCAATCT 308

QY 715 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 774  
Db 307 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 248

QY 775 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 834  
Db 247 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 188

QY 835 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTA 894  
Db 187 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTA 128

QY 895 TAAATTTGAACAGTTGCTTAATAACGTTTCTTTAATTAATAAATAAGAACTTGACTA 1006  
Db 127 TAAATTTGAACAGTTGCTTAATAACGTTTCTTTAATTAATAAATAAGAACTTGACTA 68

QY 955 ATAAATGAATGACCAAGACTTTTAAATTAATAAATAAATAAGAACTTGACTA 1006  
Db 67 ATAAATGAATGACCAAGACTTTTAAATTAATAAATAAATAAGAACTTGACTA 16

RESULT 5  
US-09-352-616A-206/c  
; Sequence 206, Application US/09352616A  
; Patent No. 6395278  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Harlocker, Susan Louise  
; APPLICANT: Jiang, Yuqui  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Mitcham, Jennifer Lynn  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; FILE REFERENCE: 210121.427C8  
; CURRENT APPLICATION NUMBER: US/09/352,616A  
; CURRENT FILING DATE: 1999-07-13  
; NUMBER OF SEQ ID NOS: 472  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 206  
; LENGTH: 487  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(487)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-352-616A-206

Query Match 37.4%; Score 439.6; DB 4; Length 487;  
Best Local Similarity 96.0%; Pred. No. 2.3e-101;  
Matches 453; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 536 TTTTGAACAAGTGCACCAAGATGCCACAAATGTGGCAGAGCCTTTGAGGAAGCGGTTCG 595  
Db 487 TTTTGAACAAGTGCACCAAGATGCCACAAATGTGGCAGAGCCTTTGAGGAAGCGGTTCG 428

QY 596 AAGAGTCTTCTACCGAGGATAGTGCAGATCATTTGATTTAGATTGTTGATGCAATCT 655  
Db 427 AAGAGTCTTCTACCGAGGATAGTGCAGATCATTTGATTTAGATTGTTGATGCAATCT 368

QY 656 TCACCGAAAGCCCAAGCCTAGC-TCATCTTCTGCTGTGATTTAGATTGTTGATGCAATCT 714  
Db 367 TCACCGAAAGCCCAAGCCTAGCCTTCTGCTGTGATTTAGATTGTTGATGCAATCT 308

QY 715 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 774  
Db 307 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 248

QY 775 AAGAGTCTTCTACCGAGGATAGTGCAGATCATTTGATTTAGATTGTTGATGCAATCT 655

Db 427 AAGAGTCTTCTACCGAGGATAGTGCAGATCATTTGATTTAGATTGTTGATGCAATCT 368  
QY 656 TCACCGAAAGCCCAAGCCTAGC-TCATCTTCTGCTGTGATTTAGATTGTTGATGCAATCT 714  
Db 367 TCACCGAAAGCCCAAGCCTAGCCTTCTGCTGTGATTTAGATTGTTGATGCAATCT 308

QY 715 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 774  
Db 307 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 248

QY 775 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 834  
Db 247 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 188

QY 835 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTA 894  
Db 187 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTA 128

QY 895 TAAATTTGAACAGTTGCTTAATAACGTTTCTTTAATTAATAAATAAGAACTTGACTA 954  
Db 127 TAAATTTGAACAGTTGCTTAATAACGTTTCTTTAATTAATAAATAAGAACTTGACTA 68

QY 955 ATAAATGAATGACCAAGACTTTTAAATTAATAAATAAATAAGAACTTGACTA 1006  
Db 67 ATAAATGAATGACCAAGACTTTTAAATTAATAAATAAATAAGAACTTGACTA 16

RESULT 6  
US-09-232-149A-206/c  
; Sequence 206, Application US/09232149A  
; Patent No. 6465611  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer Lynn  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE  
; FILE REFERENCE: 210121.427C6  
; CURRENT APPLICATION NUMBER: US/09/232,149A  
; CURRENT FILING DATE: 1999-01-15  
; NUMBER OF SEQ ID NOS: 338  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 206  
; LENGTH: 487  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(487)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-232-149A-206

Query Match 37.4%; Score 439.6; DB 4; Length 487;  
Best Local Similarity 96.0%; Pred. No. 2.3e-101;  
Matches 453; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 536 TTTTGAACAAGTGCACCAAGATGCCACAAATGTGGCAGAGCCTTTGAGGAAGCGGTTCG 595  
Db 487 TTTTGAACAAGTGCACCAAGATGCCACAAATGTGGCAGAGCCTTTGAGGAAGCGGTTCG 428

QY 596 AAGAGTCTTCTACCGAGGATAGTGCAGATCATTTGATTTAGATTGTTGATGCAATCT 655  
Db 427 AAGAGTCTTCTACCGAGGATAGTGCAGATCATTTGATTTAGATTGTTGATGCAATCT 368

QY 656 TCACCGAAAGCCCAAGCCTAGC-TCATCTTCTGCTGTGATTTAGATTGTTGATGCAATCT 714  
Db 367 TCACCGAAAGCCCAAGCCTAGCCTTCTGCTGTGATTTAGATTGTTGATGCAATCT 308

QY 715 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 774  
Db 307 TAAACCAACTCACACATATACACAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTG 248

QY 775 AAGAGTCTTCTACCGAGGATAGTGCAGATCATTTGATTTAGATTGTTGATGCAATCT 655



QY 596 AAGAGTTCTTGTACCGAGGATAGTCAGATCAATTTGATTGATCAGACAGACAGTCATCT 655  
Db 416 AAGAGTTCTTGTCCACCGAGGATAGTCAGATCAATTTGATTGATCAGACAGACAGTCATTT 357  
QY 656 TCACCGAAAGCCCAAGCCCTAGCTCATCTTGTCTGTGTTGATTGTTAGATTGTTGATGCAATCT 715  
Db 356 TCACCGAAAGCCCAAGCCCTAGCTCATCTTGTCTGTGTTGATTGTTAGATTGTTGATGCAATCT 299  
QY 716 AACCAACTCACATATACACAAATCAACATCGGGATGGAGAGAAATTAGCGTTTGC 775  
Db 298 -ACCAACTCACATATACACAAATATC-ACATGGGGATGGAGAGAAATTAGCGTTTGC 241  
QY 776 AGCAGTGTATCATCTACTATAATAAATAAATTAACCTTAATGTTGCTCTTCAATTTAGTTGGTGGGA 835  
Db 240 AGCAGTGTAT-ATCTACTATAATAAATAAATTAACCTTAATGTTGCTG-TTCATTAGTTGGTGGGA 183  
QY 836 GAAGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGACACCTTACATTTAT 895  
Db 182 GAAGGACACATCCACTCTTGGAGGAATATA-TTACTCAATAATGACACCTTACATTTAT 124  
QY 896 AATTTGTAACAGTCTCTTAATAACCTTTCTTTAAATTAATAATGTAAGTTGCAGAGCTAA 955  
Db 123 AATTTGT-ACAGTTGCTTAA-AACGTTTCTTTAAATTAATAATGTAAGTTGCAGAGCTAA 66  
QY 956 TAAATGAATGACCAAGACTTTTAAATTAATAATAAATAAAGAACTTGACTA 1006  
Db 65 TAAATGAATGACCAAGACTTTTAAATTAATAATAAATAAAGAACTTGACTA 16

## RESULT 9

US-09-439-313-106/c  
; Sequence 106, Application US/09439313  
; Patent No. 6329505  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiaangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan Louise  
; APPLICANT: Jiang Yuqi  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Kalos, Michael  
; APPLICANT: Fanger, Gary  
; APPLICANT: Retter, Mark  
; APPLICANT: Solk, John  
; APPLICANT: Day, Craig  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
; FILE REFERENCE: 210121.427C9  
; CURRENT APPLICATION NUMBER: US/09/439,313  
; CURRENT FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 575  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 106  
; LENGTH: 473  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-439-313-106

Query Match 26.1%; Score 306.2; DB 4; Length 473;  
Best Local Similarity 93.4%; Pred. No. 7.1e-68;  
Matches 440; Conservative 0; Mismatches 18; Indels 13; Gaps 11;  
QY 536 TTTTGAACACAGTGCACAAAGATGCCACAAATGGCAGCAGCCTTTGAGGAGCGGTTGC 595  
Db 473 TTTTGAACACAGTGCACAA--GAGCCACAAATGGCAGC-GCCTTTGAGGAGCGGTTGC 417  
QY 596 AAGAGTTCTTGTACCGAGGATAGTCAGATCAATTTGATTGATCAGACAGACAGTCATCT 655  
Db 416 AAGAGTTCTTGTCCACCGAGGATAGTCAGATCAATTTGATTGATCAGACAGACAGTCATTT 357  
QY 656 TCACCGAAAGCCCAAGCCCTAGCTCATCTTGTCTGTGTTGATTGTTAGATTGTTGATGCAATCT 715

Db 356 TCACCGAAAGCCCAAGCCCTAGCTCATCTTGTCTGTGTTGATTGT--AGATTGTGATGCAATCT 299  
QY 716 AACCAACTCACATATACACAAATCAACATCGGGATGGAGAGAAATTAGCGTTTGC 775  
Db 298 -ACCAACTCACATATACACAAATATC-ACATGGGGATGGAGAGAAATTAGCGTTTGC 241  
QY 776 AGCAGTGTATCATCTACTATAATAAATAAATTAACCTTAATGTTGCTCTTCAATTTAGTTGGTGGGA 835  
Db 240 AGCAGTGTAT-ATCTACTATAATAAATAAATTAACCTTAATGTTGCTG-TTCATTAGTTGGTGGGA 183  
QY 836 GAAGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGACACCTTACATTTAT 895  
Db 182 GAAGGACACATCCACTCTTGGAGGAATATA-TTACTCAATAATGACACCTTACATTTAT 124  
QY 896 AATTTGTAACAGTCTCTTAATAACCTTTCTTTAAATTAATAATGTAAGTTGCAGAGCTAA 955  
Db 123 AATTTGT-ACAGTTGCTTAA-AACGTTTCTTTAAATTAATAATGTAAGTTGCAGAGCTAA 66  
QY 956 TAAATGAATGACCAAGACTTTTAAATTAATAATAAATAAAGAACTTGACTA 1006  
Db 65 TAAATGAATGACCAAGACTTTTAAATTAATAATAAATAAAGAACTTGACTA 16

## RESULT 10

US-09-352-616A-106/c  
; Sequence 106, Application US/09352616A  
; Patent No. 6395278  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Harlocker, Susan Louise  
; APPLICANT: Jiang, Yuqi  
; APPLICANT: Xu, Jiaangchun  
; APPLICANT: Mitcham, Jennifer Lynn  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; FILE REFERENCE: 210121.427C8  
; CURRENT APPLICATION NUMBER: US/09/352,616A  
; CURRENT FILING DATE: 1999-07-13  
; NUMBER OF SEQ ID NOS: 472  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 106  
; LENGTH: 473  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-352-616A-106

Query Match 26.1%; Score 306.2; DB 4; Length 473;  
Best Local Similarity 93.4%; Pred. No. 7.1e-68;  
Matches 440; Conservative 0; Mismatches 18; Indels 13; Gaps 11;  
QY 536 TTTTGAACACAGTGCACAAAGATGCCACAAATGGCAGCAGCCTTTGAGGAGCGGTTGC 595  
Db 473 TTTTGAACACAGTGCACAA--GAGCCACAAATGGCAGC-GCCTTTGAGGAGCGGTTGC 417  
QY 596 AAGAGTTCTTGTACCGAGGATAGTCAGATCAATTTGATTGATCAGACAGACAGTCATCT 655  
Db 416 AAGAGTTCTTGTCCACCGAGGATAGTCAGATCAATTTGATTGATCAGACAGACAGTCATTT 357  
QY 656 TCACCGAAAGCCCAAGCCCTAGCTCATCTTGTCTGTGTTGATTGTTAGATTGTTGATGCAATCT 715  
Db 356 TCACCGAAAGCCCAAGCCCTAGCTCATCTTGTCTGTGTTGATTGT--AGATTGTGATGCAATCT 299  
QY 716 AACCAACTCACATATACACAAATCAACATCGGGATGGAGAGAAATTAGCGTTTGC 775  
Db 298 -ACCAACTCACATATACACAAATATC-ACATGGGGATGGAGAGAAATTAGCGTTTGC 241  
QY 776 AGCAGTGTATCATCTACTATAATAAATAAATTAACCTTAATGTTGCTCTTCAATTTAGTTGGTGGGA 835  
Db 240 AGCAGTGTAT-ATCTACTATAATAAATAAATTAACCTTAATGTTGCTG-TTCATTAGTTGGTGGGA 183  
QY 836 GAAGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGACACCTTACATTTAT 895  
Db 182 GAAGGACACATCCACTCTTGGAGGAATATA-TTACTCAATAATGACACCTTACATTTAT 124





Best Local Similarity 53.5%; Pred. No. 2.7e-11;  
Matches 170; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 106 TTTTAAAGTAATCTCTCGAGATGGTGGAGAGTTCACCTTATGACAGAT 165  
Db 82 TATTCAGATTACTCTGATTTGGCACTCAGGGGTGGAAAGTCTTGCTTCTTTAGGT 141  
QY 166 ATGTAACATAAATAGTTTATGATACCCAGCTCTTCCATACAAATAGGTGTGGAATTTTAAATA 225  
Db 142 TTGCAGATGATACATATACAGAAAGCTACATCAGCAGATTTGTGGAATTTCAAATAA 201  
QY 226 AAGATTGGAGTGGATGACATTTTGTATCCATGCGAGATTTGGGACACGGCAGGTGAGG 285  
Db 202 GAATCATAGATTAGACGGGAAACAAATCAAGCTTCAATATATGGGACACAGCAGGCCAGG 261  
QY 286 AGCATTCGGAAGCTCGAGGACACCATTTTACAGAGTTTGTACTGTGCTGCTTACTT 345  
Db 262 AAGATTTCGAACATCACCTCCAGTTATTACAGAGAGCCCATGGATCATATGTTGTGT 321  
QY 346 TTAGTGTGATGATTCACAAAGCTTCCAGAACTTAACTGGAAGAAAGAAATTCATAT 405  
Db 322 ATGATGTGACAGATCAGAGTCCCTTCAATAATGTTAAACAGTGGCTGCAGGAAATAGATC 381  
QY 406 ATTATGAGATGTGAAAG 423  
Db 382 GTTATGCCAGTGAATG 399

RESULT 15

US-09-016-434-1060  
; Sequence 1060, Application US/09016434  
; Patent No. 6500938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,434  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0002 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1060:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 944 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENBANK

CLONE: g1055280  
US-09-016-434-1060

Query Match 5.9%; Score 69.6; DB 4; Length 944;  
Best Local Similarity 53.4%; Pred. No. 2.4e-08;  
Matches 175; Conservative 0; Mismatches 144; Indels 9; Gaps 1;

QY 257 CATGCAGATTTGGGACACGGCAGGTTCAGAGCGATTCCGAGCCTTGAGACACCACTTTTA 316  
Db 452 CTTGAGTTTATGGGACACAGCGGCGAGGAGGTTTCGTAGCTTAACGACAGCGCTTTT 511  
QY 317 CAGAGGTTCTGACTGCTGCTGCTTACTTTTGTAGTGTGATGATTCAAAAGCTTCCAGAA 376  
Db 512 CAGAGATGCTATGGGTTTCTTCTTACTTTGTGATCTGACAAATGAGCAAGTTTCTCAA 571  
QY 377 CTTAAGTAACTGGAAGAAGAAATTCATATATATGACAGATGTGAAAGCCTTGAGAGCTT 436  
Db 572 TGTCAAGAACTGGATATAGCCAGCTACAGATGCTATGTCATATTTGTGAAAACCCAGA----- 625  
QY 437 TCCTTTTGTGATTTCTGGGTAAACAAGATTGACATAAGCGAAGCGGAGGTGTCTACAGAGA 496  
Db 626 ---TATAGTGTGCTGGAAACAGAGTGTCTGGAGGACCGAGAGTAGTGAAGAGGA 682  
QY 497 AGCCCAAGCTTGTGTCAGGGCAACAGCGGCACTATCCTTTATTTTGAACAAGTGCAGAA 556  
Db 683 GGAAGCCATAGCACTTCGCGAGAGAAATATGGAATCCCTTACTTTGAAACTAGTGTGCTG 742  
QY 557 TGCCACAAATGTGGCAGCAGCGCTTTGAG 584  
Db 743 TGGGACAAACATAGCCCAAGCAATTGAG 770

Search completed: February 10, 2004, 04:46:01  
Job time : 97 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2004, 02:32:31 ; Search time 489 Seconds

(without alignments)  
8851.265 Million cell updates/sec

Title: US-09-988-974-6

Perfect score: 1175

Sequence: 1 CACTGTGATGAACACTTTT.....GTAAAAA.....1175

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata1/pubpna/PCTU5\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata1/pubpna/US10\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata1/pubpna/US60\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1175	100.0	1175	9	US-09-988-974-6
2	1160.2	98.7	1372	13	Sequence 6, Appli
3	1100	93.6	1969	9	Sequence 488, App
4	1100	93.6	6058	11	Sequence 10308, A
5	632	53.8	632	9	Sequence 5493, Ap
6	604.4	51.4	606	12	Sequence 26942, A
7	439.6	37.4	487	9	Sequence 1082, Ap
8	439.6	37.4	487	9	Sequence 206, App
9	439.6	37.4	487	9	Sequence 206, App
10	439.6	37.4	487	9	Sequence 206, App
11	439.6	37.4	487	9	Sequence 206, App
12	439.6	37.4	487	9	Sequence 206, App
13	439.6	37.4	487	10	Sequence 206, App
14	439.6	37.4	487	10	Sequence 206, App
15	439.6	37.4	487	13	Sequence 206, App

c 16	439.6	37.4	487	13	US-10-294-025-206	Sequence 206, App
c 17	439.6	37.4	487	14	US-10-012-896-206	Sequence 206, App
c 18	439.6	37.4	487	15	US-10-010-940-206	Sequence 206, App
c 19	394.4	33.6	397	12	US-10-242-535A-42720	Sequence 42720, A
c 20	356	30.3	393	13	US-10-101-510-138	Sequence 138, App
c 21	333.2	28.4	648	13	US-10-029-386-24799	Sequence 24799, A
c 22	333.2	28.4	1114	10	US-09-764-868-508	Sequence 508, App
c 23	332.4	27.4	1111	10	US-09-764-868-89	Sequence 89, Appli
c 24	306.2	26.1	473	9	US-09-759-143-106	Sequence 106, App
c 25	306.2	26.1	473	9	US-09-780-669-106	Sequence 106, App
c 26	306.2	26.1	473	9	US-09-030-606-106	Sequence 106, App
c 27	306.2	26.1	473	9	US-09-822-827-106	Sequence 106, App
c 28	306.2	26.1	473	9	US-09-115-453-106	Sequence 106, App
c 29	306.2	26.1	473	10	US-09-232-880-106	Sequence 106, App
c 30	306.2	26.1	473	10	US-09-895-793-106	Sequence 106, App
c 31	306.2	26.1	473	10	US-09-895-813-106	Sequence 106, App
c 32	306.2	26.1	473	13	US-10-144-678A-106	Sequence 106, App
c 33	306.2	26.1	473	13	US-10-294-025-106	Sequence 106, App
c 34	306.2	26.1	473	14	US-10-012-896-106	Sequence 106, App
c 35	306.2	26.1	473	15	US-10-010-940-106	Sequence 106, App
c 36	188.4	16.0	210	12	US-10-242-535A-93	Sequence 93, Appli
c 37	185.8	15.8	462	9	US-09-864-761-2267	Sequence 2267, Ap
c 38	183.4	15.6	2178	9	US-09-925-302-186	Sequence 186, App
c 39	183.4	15.6	2302	12	US-10-264-049-724	Sequence 724, App
c 40	181.8	15.5	624	12	US-10-305-720-1083	Sequence 1083, Ap
c 41	141.2	12.0	627	12	US-10-369-493-25550	Sequence 25550, A
c 42	137.4	11.7	698	9	US-09-828-310-8	Sequence 8, Appli
c 43	133.4	11.4	223	9	US-09-864-761-19005	Sequence 19005, A
c 44	128.4	10.9	941	9	US-09-770-445-358	Sequence 358, App
c 45	128.2	10.9	612	10	US-09-938-842A-1196	Sequence 1196, Ap

ALIGNMENTS

RESULT 1  
US-09-988-974-6  
; Sequence 6, Application US/09988974  
; Patent No. US20020090712A1  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; Goli, Surya K.  
; Bandman, Olga  
; TITLE OF INVENTION: NOVEL RAB PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: fastseq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09988,974  
; FILING DATE: 19-No. US20020090712A1-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/215,887  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0183 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1175 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Consensus  
CLONE: Consensus  
SEQUENCE DESCRIPTION: SEQ ID NO: 6;  
US-09-988-974-6

Query Match 100.0%; Score 1175; DB 9; Length 1175;  
Best Local Similarity 100.0%; Pred. No. 5.5e-253;  
Matches 1175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACTGTGATGAACACTTTCCCGTGTGCTTGGAGTGCATCTTCAACAACCCCTAGGAG 60  
DB 1 CACTGTGATGAACACTTTCCCGTGTGCTTGGAGTGCATCTTCAACAACCCCTAGGAG 60  
QY 61 GGTCTCTGAAGCTTTTGAGATTAAACAATGGCAGGAAATCATCACTTTTAAAGTAATTC 120  
DB 61 GGTCTCTGAAGCTTTTGAGATTAAACAATGGCAGGAAATCATCACTTTTAAAGTAATTC 120  
QY 121 TCCTTGGAGATGGTGGAGTGGGAAGTTCATTAAGACAGATATCTAATAAGT 180  
DB 121 TCCTTGGAGATGGTGGAGTGGGAAGTTCATTAAGACAGATATCTAATAAGT 180  
QY 181 TTGATACCCAGCTTCCATCAATAGTGTGGAAATTTTAAATAAGATTTGGAAGTGG 240  
DB 181 TTGATACCCAGCTTCCATCAATAGTGTGGAAATTTTAAATAAGATTTGGAAGTGG 240  
QY 241 ATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTTCAGGAGCC 300  
DB 241 ATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTTCAGGAGCC 300  
QY 301 TGAGGACACATTTACAGAGTTCATCTGCTGCTGCTTACCTTTTAGTTCGATGATT 360  
DB 301 TGAGGACACATTTACAGAGTTCATCTGCTGCTGCTTACCTTTTAGTTCGATGATT 360  
QY 361 CACAAAGCTTCAGAACCTTAACTGGAAGAAAGATTCATATATTATGACAGATGTA 420  
DB 361 CACAAAGCTTCAGAACCTTAACTGGAAGAAAGATTCATATATTATGACAGATGTA 420  
QY 421 AAGAGCTGAGAGCTTCCCTTTTGTGATCTCTGCTGCTGCTTACCTTTTAGTTCGATGATT 480  
DB 421 AAGAGCTGAGAGCTTCCCTTTTGTGATCTCTGCTGCTGCTTACCTTTTAGTTCGATGATT 480  
QY 481 AGGTGCTACAGAGAGCCCAAGCTTGGTGCAGGACAAACGGCAGCTATCTTATTG 540  
DB 481 AGGTGCTACAGAGAGCCCAAGCTTGGTGCAGGACAAACGGCAGCTATCTTATTG 540  
QY 541 AAACAAGTCCAAAGATGCCAATAATGTGGCAGCAGCCCTTTGAGGAAGCGGTTCGAAGAG 600  
DB 541 AAACAAGTCCAAAGATGCCAATAATGTGGCAGCAGCCCTTTGAGGAAGCGGTTCGAAGAG 600  
QY 601 TTCTTGCTACCGAGGATAGTGCATCTATTCATTCAGACACACAGTCAATCTTACC 660  
DB 601 TTCTTGCTACCGAGGATAGTGCATCTATTCATTCAGACACACAGTCAATCTTACC 660  
QY 661 GAAAGCCCAAGCTAGCTCATCTTGTGTTGATTGTTAGATTGTTGATGATTCATTAACCA 720  
DB 661 GAAAGCCCAAGCTAGCTCATCTTGTGTTGATTGTTAGATTGTTGATGATTCATTAACCA 720  
QY 721 ACTCACATATACACAAATCAACATGGGGATGGAGAGAAATAGCGTTTCAGGAG 780  
DB 721 ACTCACATATACACAAATCAACATGGGGATGGAGAGAAATAGCGTTTCAGGAG 780  
QY 781 TGTATCATCTACTAATAAATTAATTAATGTGCTGCTTCAATAGTTGGTGGGAGAGG 840  
DB 781 TGTATCATCTACTAATAAATTAATTAATGTGCTGCTTCAATAGTTGGTGGGAGAGG 840  
QY 841 GACACATCCACTCTTGGAGGAATATATTACTCAATATGACACCTTCATTTAAAT 900  
DB 841 GACACATCCACTCTTGGAGGAATATATTACTCAATATGACACCTTCATTTAAAT 900

DB 841 GACACATCCACTCTTGGAGGAATATATTACTCAATATGACACCTTCATTTAAAT 900  
QY 901 GTAAACAGTTGTCTAATAACGTTTCTTTAATTTAAATATGTAAGTTGCAGAGCTAATAAT 960  
DB 901 GTAAACAGTTGTCTAATAACGTTTCTTTAATTTAAATATGTAAGTTGCAGAGCTAATAAT 960  
QY 961 GAAATGACCAAGACTTTTAATTAATAAATAAAGAACTTGCATATTCTAGAAGTTATA 1020  
DB 961 GAAATGACCAAGACTTTTAATTAATAAATAAAGAACTTGCATATTCTAGAAGTTATA 1020  
QY 1021 CTTGGATTTTCTCGGAAAAATGGAACTACTTTTTATATGTGTATGTTTTATGCAA 1080  
DB 1021 CTTGGATTTTCTCGGAAAAATGGAACTACTTTTTATATGTGTATGTTTTATGCAA 1080  
QY 1081 TTAGCATTTCTATTCTTGGTTCAGGAAATACCTTCTTAAGCAATAATCTTAGATATTA 1140  
DB 1081 TTAGCATTTCTATTCTTGGTTCAGGAAATACCTTCTTAAGCAATAATCTTAGATATTA 1140  
QY 1141 AGATTAAATCTAATGTAATAAATAAATAAATAAATAA 1175  
DB 1141 AGATTAAATCTAATGTAATAAATAAATAAATAAATAA 1175

RESULT 2  
US-10-101-510-488  
; Sequence 488, Application US/10101510  
; Publication No. US20030148295A1  
; GENERAL INFORMATION:  
; APPLICANT: WAN, JACKSON  
; APPLICANT: WANG, YIXIN  
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE  
; FILE REFERENCE: 15117.0012  
; CURRENT APPLICATION NUMBER: US/10/101,510  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/276,947  
; PRIOR FILING DATE: 2001-03-20  
; NUMBER OF SEQ ID NOS: 805  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 488  
; LENGTH: 1372  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (1362)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (1364)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (1366)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (1368)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-101-510-488

Query Match 98.7%; Score 1160.2; DB 13; Length 1372;  
Best Local Similarity 99.2%; Pred. No. 1.2e-249;  
Matches 1165; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CACTGTGATGAACACTTTCCCGTGTGCTTGGAGTGCATCTTCAACAACCCCTAGGAG 60  
DB 190 CACTGTGATGAACACTTTCCCGTGTGCTTGGAGTGCATCTTCAACAACCCCTAGGAG 249  
QY 61 GGTCTCTGAAGCTTTTGAGATTAAACAATGGCAGGAAATCATCACTTTTAAAGTAATTC 120  
DB 250 GGTCTCTGAAGCTTTTGAGATTAAACAATGGCAGGAAATCATCACTTTTAAAGTAATTC 309  
QY 121 TCCTTGGAGATGGTGGAGTGGGAAGTTCATTTATGAAACAGATATGTAATAAGT 180



```

310 TCCTGGAGATGGTGGAGTTCGGAGAGTTCACATTATGAACAGATATGAACATTAAGT 369
181 TTGATACCCAGCTCTCCATCAATAGGTGGGATTTTAAATAAAGATTTGGAAGTGG 240
370 TTGATACCCAGCTCTCCATCAATAGGTGGGATTTTAAATAAAGATTTGGAAGTGG 429
241 ATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTTCAGGAGCGATTCGGAAGCC 300
430 ATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTTCAGGAGCGATTCGGAAGCC 489
301 TGAGACACCATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAGTGTGATGATT 360
490 TGAGACACCATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAGTGTGATGATT 549
361 CACAAGCTTCCAGAACTTAAGTAACCTGGAAGAAAGAAATCATATATATGCGAGATGTA 420
550 CACAAGCTTCCAGAACTTAAGTAACCTGGAAGAAAGAAATCATATATATGCGAGATGTA 609
421 AAGAGCTGAGAGCTTTCCTTTTGTGATCTGGGTAAACAAGATTGACATAAGCGAACGGC 480
610 AAGAGCTGAGAGCTTTCCTTTTGTGATCTGGGTAAACAAGATTGACATAAGCGAACGGC 669
481 AGGTGCTACAGAGAGCCAGCTTGTGGGACGAGCGGCTTGGAGGAGCGGTTCTTATTTTG 540
670 AGGTGCTACAGAGAGCCAGCTTGTGGGACGAGCGGCTTGGAGGAGCGGTTCTTATTTTG 729
541 AAACAAGTGAAGAAGATGCGCAAAATGTCGAGCAGCGCTTGGAGGAGCGGTTCTTATTTTG 600
730 AAACAAGTGAAGAAGATGCGCAAAATGTCGAGCAGCGCTTGGAGGAGCGGTTCTTATTTTG 789
601 TTCTTGTACCGAGGATAGTCAGATCAATTTGATTCAGACAGACAGATCAATCTTCACC 660
790 TTCTTGTACCGAGGATAGTCAGATCAATTTGATTCAGACAGACAGATCAATCTTCACC 849
661 GAAAGCCCAAGCTAGCTCATCTGCTGTTGATTTGTAGATTCTGATGATCTCAACCA 720
850 GAAAGCCCAAGCTAGCTCATCTGCTGTTGATTTGTAGATTCTGATGATCTCAACCA 909
721 ACTCACACATATACAAAATCAACATGCGGATGGAGAGAGAAATAGCGTTTCAGACAG 780
910 ACTCACACATATACAAAATCAACATGCGGATGGAGAGAGAAATAGCGTTTCAGACAG 969
781 TGTATCATCTACTAATAAATTAACCTAATGTTGCTGCTTCAATAGTTTGGTGGGAGAG 840
970 TGTATCATCTACTAATAAATTAACCTAATGTTGCTGCTTCAATAGTTTGGTGGGAGAG 1029
841 GACACATCCACTCTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTATAAAT 900
1030 GACACATCCACTCTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTATAAAT 1089
901 GTAACAGTTGCTAATAACGTTTCTTTAATTAATTAATGATGATGATGATGATTAAT 960
1090 GTAACAGTTGCTAATAACGTTTCTTTAATTAATTAATGATGATGATGATTAAT 1149
961 GAAATGACCAAGACTTTAATAATAAATAAAGAACTTGAATTTCTAGAAGTTATA 1020
1150 GAAATGACCAAGACTTTAATAATAAATAAAGAACTTGAATTTCTAGAAGTTATA 1209
1021 CTGGATTTTTCCTGGGAAATGAGAACTATTTTATATGATGATGATGATGATGATGAT 1080
1210 CTGGATTTTTCCTGGGAAATGAGAACTATTTTATATGATGATGATGATGATGATGAT 1269
1081 TTAGCATTTGATTTCTGGTTTCAGGAAATATCTTCTTAAGCAATTAATGATGATTTAA 1140
1270 TTAGCATTTGATTTCTGGTTTCAGGAAATATCTTCTTAAGCAATTAATGATGATTTAA 1329
1141 AGATTAATTAATCTAATGATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1174
1330 AGATTAATTAATCTAATGATTTGATTAATAAATAAATAAATAAATAAATAAATAAATAA 1363

```

RESULT 3

```

US-09-864-761-10308
; Sequence 10308, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10308
; LENGTH: 1969
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC003037.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
US-09-864-761-10308

```

```

Query Match 93.6%; Score 1100; DB 9; Length 1969;
Best Local Similarity 100.0%; Pred. No. 4.4e-236;
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 AGGTTCTTGAAGCTTTTGAGATTACAAATGGCAGGAATCATCACTTTTAAAGTAAT 118
DB 311 AGGTTCTTGAAGCTTTTGAGATTACAAATGGCAGGAATCATCACTTTTAAAGTAAT 370
QY 119 TCTCTTGGAGATGGTGGAGTTCGGAGAGTTCACCTTATGAAACAGATATGTAACATAA 178

```

371 TCCTCTGGAGATGGTGGAGTGGGAGAGTTCATTTATGACAGATATGTAATTAATAA 430  
 179 GTTTGATACCCAGCTCTCCATACAAATAGTGTGGAAATTTTAAATAAGAGTTTGAAGT 238  
 431 GTTTGATACCCAGCTCTCCATACAAATAGTGTGGAAATTTTAAATAAGAGTTTGAAGT 490  
 239 GGATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTCAGAGCATTCGGAAG 298  
 491 GGATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTCAGAGCATTCGGAAG 550  
 299 CTTGAGGACACCATTTTACAGAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358  
 551 CTTGAGGACACCATTTTACAGAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 610  
 359 TTCAAAAGCTTCCAGAACTTAACTAGTCTGGAAGAGATTCATATATATGACAGATGT 418  
 611 TTCAAAAGCTTCCAGAACTTAACTAGTCTGGAAGAGATTCATATATATGACAGATGT 670  
 419 GAAAGAGCTTGAGAGCTTTCTTTTGTGATTTCTGGGTAAACAAGATTCACATAGCGAAG 478  
 671 GAAAGAGCTTGAGAGCTTTCTTTTGTGATTTCTGGGTAAACAAGATTCACATAGCGAAG 730  
 479 GCAGGTGTCTACAGAGAACCCAGCTTGGTGCAGGACACCGGACATATCTTTATTT 538  
 731 GCAGGTGTCTACAGAGAACCCAGCTTGGTGCAGGACACCGGACATATCTTTATTT 790  
 539 TGAACAAGTGCAGAAAGATCCCAAAATGTGGCAGCAGCTTTGAGGAGAGCGGTTGGAAG 598  
 791 TGAACAAGTGCAGAAAGATCCCAAAATGTGGCAGCAGCTTTGAGGAGAGCGGTTGGAAG 850  
 599 AGTTCTGCTACCGAGATAGGTGAGATCAATTTGATTCAGACAGACAGTCAATCTTCA 658  
 851 AGTTCTGCTACCGAGATAGGTGAGATCAATTTGATTCAGACAGACAGTCAATCTTCA 910  
 659 CCGAAGCCCAAGCTAGCTCATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 718  
 911 CCGAAGCCCAAGCTAGCTCATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 970  
 719 CAATCACACATATACAAAATCAACATGGGATGGAGAGAGATTTAGCGTTTGCAGC 778  
 971 CAATCACACATATACAAAATCAACATGGGATGGAGAGAGATTTAGCGTTTGCAGC 1030  
 779 AGGTATCATCTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 838  
 1031 AGGTATCATCTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1090  
 839 GGGACATCCACTCTTGGAGGATATATTTACTCAATAATGGCACCCTTACATTTATATAA 898  
 1091 GGGACATCCACTCTTGGAGGATATATTTACTCAATAATGGCACCCTTACATTTATATAA 1150  
 899 TTGTAACAGTTGTCTAATAACGTTTCTTTAATTAATATGTAAGTTGAGAGCTTAATAA 958  
 1151 TTGTAACAGTTGTCTAATAACGTTTCTTTAATTAATATGTAAGTTGAGAGCTTAATAA 1210  
 959 ATGAAATGACCAAGACTTTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1018  
 1211 ATGAAATGACCAAGACTTTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1270  
 1019 TACTTGGAATTTTCTGGGAAATGGAGACATPACTTTTTATATGATGATGTTTTATGTC 1078  
 1271 TACTTGGAATTTTCTGGGAAATGGAGACATPACTTTTTATATGATGATGTTTTATGTC 1330  
 1079 AATTAGCATTTGATTTCTGGTTCAGGAAATACCTTTCCTAAGCAATATGATGATGATTT 1138  
 1331 AATTAGCATTTGATTTCTGGTTCAGGAAATACCTTTCCTAAGCAATATGATGATGATTT 1390  
 1139 AAAGATTAATAATCTAATGTA 1158  
 1391 AAAGATTAATAATCTAATGTA 1410

RESULT 4

US-09-764-891-5493  
 ; Sequence 5493, Application US/09764891  
 ; Publication No. US20030077808A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PC006  
 ; CURRENT APPLICATION NUMBER: US/09/764,891  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 10231  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 5493  
 ; LENGTH: 6058  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-764-891-5493

Query Match 93.6%; Score 1100; DB 11; Length 6058;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-236;  
 Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	59	AGGTTCTTGAAGCTTTTGGAGATTAA	CAATGGCGAGGAAATCATCTTTTAAAGTAAT	118
Db	4905	AGGTTCTTGAAGCTTTTGGAGATTAA	CAATGGCGAGGAAATCATCTTTTAAAGTAAT	4964
Qy	119	TCTCTTGGAGATGGTGGAGATTGGA	AGATTGCAACAGATATGTAATAATAA	178
Db	4965	TCTCTTGGAGATGGTGGAGATTGGA	AGATTGCAACAGATATGTAATAATAA	5024
Qy	179	GTTTGATACCCAGCTCTTCCATACAA	TAGGTGGAAATTTTAAATAAGATTGGAAGT	238
Db	5025	GTTTGATACCCAGCTCTTCCATACAA	TAGGTGGAAATTTTAAATAAGATTGGAAGT	5084
Qy	239	GGATGGACATTTTGTACCATGAGATT	TGGGACACGGCAGGTCAGAGCGATTCGGAAG	298
Db	5085	GGATGGACATTTTGTACCATGAGATT	TGGGACACGGCAGGTCAGAGCGATTCGGAAG	5144
Qy	299	CCTGAGGACACCAATTTTACAGAGGT	CTGACTGCTGCTGCTTACTTTTGTAGTGTGATGA	358
Db	5145	CCTGAGGACACCAATTTTACAGAGGT	CTGACTGCTGCTGCTTACTTTTGTAGTGTGATGA	5204
Qy	359	TTCAAAAGCTTCCAGAACTTAACTAG	TAACTGGAGAAAGAAATTCATATATATGCGAGATGT	418
Db	5205	TTCAAAAGCTTCCAGAACTTAACTAG	TAACTGGAGAAAGAAATTCATATATATGCGAGATGT	5264
Qy	419	GAAAGAGCTTGAGAGCTTTCTTTTGT	GATTCTGGGTAAACAAGATTGACATAAGCGGAACG	478
Db	5265	GAAAGAGCTTGAGAGCTTTCTTTTGT	GATTCTGGGTAAACAAGATTGACATAAGCGGAACG	5324
Qy	479	GCAGGTGTCTACAGAGAACCCCAAG	CTTGGTGGAGGGAACAACGGGACATCTCTTATTTT	538
Db	5325	GCAGGTGTCTACAGAGAACCCCAAG	CTTGGTGGAGGGAACAACGGGACATCTCTTATTTT	5384
Qy	539	TGAAACAAGTGCAGAAAGATGCCA	AAATGTCAGACAGCCTTTGAGGAGCGGTTGGAAG	598
Db	5385	TGAAACAAGTGCAGAAAGATGCCA	AAATGTCAGACAGCCTTTGAGGAGCGGTTGGAAG	5444
Qy	599	AGTTCTTGTCTACCGAGGATAGGT	CAGATCATCTTTGATTTCAGACAGACACAGTCTTCTCA	658
Db	5445	AGTTCTTGTCTACCGAGGATAGGT	CAGATCATCTTTGATTTCAGACAGACACAGTCTTCTCA	5504
Qy	659	CCGAAAGCCCAAGCTAGCTCATCTT	GCTGCTGCTTGTAGATTGTTGATGTCATCTTAAC	718
Db	5505	CCGAAAGCCCAAGCTAGCTCATCTT	GCTGCTGCTTGTAGATTGTTGATGTCATCTTAAC	5564
Qy	719	CAACTCACATATACAAAAATCAACA	TGGGATGGAGAGAGAAATGACGTTTTCGAGC	778
Db	5565	CAACTCACATATACAAAAATCAACA	TGGGATGGAGAGAGAAATGACGTTTTCGAGC	5624
Qy	779	AGTGTATCATCTACTAATAAATAA	ATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	838
Db	5625	AGTGTATCATCTACTAATAAATAA	ATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	5684

QY 839 GGGACATCCACTCTTGGAGGAATATATTACTCAATATGCACTTACCTTATATAA 898  
Db 5685 GGGACATCCACTCTTGGAGGAATATATTACTCAATATGCACTTACCTTATATAA 5744  
QY 899 TTGTAACAGTGTCTAATAACGTTCTTTAATTAAATATGTAAGTTGCAGAGCTAATAA 958  
Db 5745 TTGTAACAGTGTCTAATAACGTTCTTTAATTAAATATGTAAGTTGCAGAGCTAATAA 5804  
QY 959 ATGAATGACCAAGCACTTAAATATATAAATAAGAACTGCACTATCTAGAGTTA 1018  
Db 5805 ATGAATGACCAAGCACTTAAATATATAAATAAGAACTGCACTATCTAGAGTTA 5864  
QY 1019 TACTTGGATTTTTCCTGGGAAATGAGAACTACTTTTATATGTAATGTTTATGTC 1078  
Db 5865 TACTTGGATTTTTCCTGGGAAATGAGAACTACTTTTATATGTAATGTTTATGTC 5924  
QY 1079 AATTAGCAATGTAATCTTCTGTTTCAGGAAATACCTTCTTAAAGCAATAATGTTAGATATT 1138  
Db 5925 AATTAGCAATGTAATCTTCTGTTTCAGGAAATACCTTCTTAAAGCAATAATGTTAGATATT 5984  
QY 1139 AAGATTAAATCTAATGTA 1158  
Db 5985 AAGATTAAATCTAATGTA 6004

## RESULT 5

US-09-864-761-26942  
; Sequence 26942, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Hanzel, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Reomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 26942  
; LENGTH: 632  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC003037.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2  
; OTHER INFORMATION: EST HUMAN HIT: BE735344.1, EVALU0 0.00e+00  
; OTHER INFORMATION: NT HIT: g11421053, EVALU0 0.00e+00  
; OTHER INFORMATION: SWISSPROT HIT: P51151, EVALU0 1.00e-108  
US-09-864-761-26942

Query Match 53.8%; Score 632; DB 9; Length 632;  
Best Local Similarity 100.0%; Pred. No. 1.6e-131;  
Matches 632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 GGTCTTGAAGCTTTTGAGATTAAACAATGGCAGGAAATCATCATCTTTTAAAGTAATTC 120  
Db 1 GGTCTTGAAGCTTTTGAGATTAAACAATGGCAGGAAATCATCATCTTTTAAAGTAATTC 60  
QY 121 TCCTTGGAGATGGTGGAGTTGGGAAGAGTTCACATTATGAACAGATATGTAACTAATAGT 180  
Db 61 TCCTTGGAGATGGTGGAGTTGGGAAGAGTTCACATTATGAACAGATATGTAACTAATAGT 120  
QY 181 TTGATACCCAGCTCTTCATACAAATAGTGTGGAAATTTTAAATAAAGATTGGAAGTGG 240  
Db 121 TTGATACCCAGCTCTTCATACAAATAGTGTGGAAATTTTAAATAAAGATTGGAAGTGG 180  
QY 241 ATGGACATTTTGTACCATGCAGATTGGGACACGAGGTCAGGAGGATTCGAGAGCC 300  
Db 181 ATGGACATTTTGTACCATGCAGATTGGGACACGAGGTCAGGAGGATTCGAGAGCC 240  
QY 301 TGAGGACACCAATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAGTGTGATGATT 360  
Db 241 TGAGGACACCAATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAGTGTGATGATT 300  
QY 361 CACAAAGCTTCCAGAACTTAAAGTAACTGGAAGAAAGAAATTCATATATATGACATGTA 420  
Db 301 CACAAAGCTTCCAGAACTTAAAGTAACTGGAAGAAAGAAATTCATATATATGACATGTA 360  
QY 421 AAGAGCTTGAGAGCTTTTCTTTTGTGATTCTGGGTAAACAAGATTGACATGAAGCGAGC 480  
Db 361 AAGAGCTTGAGAGCTTTTCTTTTGTGATTCTGGGTAAACAAGATTGACATGAAGCGAGC 420  
QY 481 AGGTGCTTACAGAGAGAGCCCAAGCTTGGTCAGGAGCAACGGGACTATCTTTATTG 540  
Db 421 AGGTGCTTACAGAGAGAGCCCAAGCTTGGTCAGGAGCAACGGGACTATCTTTATTG 480  
QY 541 AAACAAGTGCAAAAGATGCCACAAATGTGGCAGCAGCTTTTGAGGAAGCGGTTCGAAGAG 600  
Db 481 AAACAAGTGCAAAAGATGCCACAAATGTGGCAGCAGCTTTTGAGGAAGCGGTTCGAAGAG 540  
QY 601 TTCTTGTACGAGGATAGTTCAGATCATTTGATTTCAGACAGACACAGTCAATCTTCAAC 660  
Db 541 TTCTTGTACGAGGATAGTTCAGATCATTTGATTTCAGACAGACACAGTCAATCTTCAAC 600  
QY 661 GAAAGCCCAAGCCCTAGCTCATCTTGTGTTGA 692  
Db 601 GAAAGCCCAAGCCCTAGCTCATCTTGTGTTGA 632

## RESULT 6

US-10-305-720-1082  
; Sequence 1082, Application US/10305720

Publication No. US20040010136A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Devin C.  
APPLICANT: Mitcham, Jennifer L.  
FILE REFERENCE: PA-0002-1 CON  
CURRENT APPLICATION NUMBER: US/10/305,720  
CURRENT FILING DATE: 2002-11-26  
PRIOR APPLICATION NUMBER: 09/016,434  
PRIOR FILING DATE: 1998-01-30  
NUMBER OF SEQ ID NOS: 1490  
SOFTWARE: PERL Program  
SEQ ID NO 1082  
LENGTH: 606  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: GenBank ID No. US20040010136A1 g1174146  
US-10-305-720-1082

Query Match 51.4%; Score 604.4; DB 12; Length 606;  
Best Local Similarity 99.8%; Pred. No. 2.3e-125; Indels 0; Gaps 0;  
Matches 605; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 87 ATGGCAGGAAATCATCATCTTTTAAAGTAATCTCTTGGAGATGGTGGAGTGGGAAG 146  
Db 1 ATGGCAGGAAATCTTCACTTTTAAAGTAATCTCTTGGAGATGGTGGAGTGGGAAG 60

QY 147 AGTTCACTTATGAACAGATGATGTAATAAGTTTGTATACCCAGCTCTTCATACATA 206  
Db 61 AGTTCACTTATGAACAGATGATGTAATAAGTTTGTATACCCAGCTCTTCATACATA 120

QY 207 GGTGTGGAATTTTAAATAAGATTTGGAAGTGGATGGACATTTTGTACCATGAGATT 266  
Db 121 GGTGTGGAATTTTAAATAAGATTTGGAAGTGGATGGACATTTTGTACCATGAGATT 180

QY 267 TGGGACACGGCAGGTGAGGAGGATCCGAAGCTTGAGGACACCATTTTACAGGTTCT 326  
Db 181 TGGGACACGGCAGGTGAGGAGGATCCGAAGCTTGAGGACACCATTTTACAGGTTCT 240

QY 327 GACTGCTGCTGCTTACTTTTATGTCGATGATTCACAAAGCTTCAGAACTTAAAGTAAC 386  
Db 241 GACTGCTGCTGCTTACTTTTATGTCGATGATTCACAAAGCTTCAGAACTTAAAGTAAC 300

QY 387 TGAAGAAAGATTCATATATATGATGAGATGGAAGAGCTTGAGAGCTTTCTTTTCTG 446  
Db 301 TGAAGAAAGATTCATATATATGATGAGATGGAAGAGCTTGAGAGCTTTCTTTTCTG 360

QY 447 ATTCTGGGTAAACAAGATTCACATAGCGAAGCGGAGGTCTTACAGAGAGCCCAAGCT 506  
Db 361 ATTCTGGGTAAACAAGATTCACATAGCGAAGCGGAGGTCTTACAGAGAGCCCAAGCT 420

QY 507 TGGTCAGGGAACAAGCGGAGCTATCTTTATTTGAAACAAGTGCAAAAGATGCCACAAT 566  
Db 421 TGGTCAGGGAACAAGCGGAGCTATCTTTATTTGAAACAAGTGCAAAAGATGCCACAAT 480

QY 567 GTGGCAGAGCTTTGAGGAAGCGGTTCGAAGTCTTGTACCGAGATAGGTGAGT 626  
Db 481 GTGGCAGAGCTTTGAGGAAGCGGTTCGAAGTCTTGTACCGAGATAGGTGAGT 540

QY 627 CATTTGATTTCAGACAGACACAGTCAATCTTACCGAAAGCCCAAGCTTAGCTCATCTTGC 686  
Db 541 CATTTGATTTCAGACAGACACAGTCAATCTTACCGAAAGCCCAAGCTTAGCTCATCTTGC 600

QY 687 TGTGGA 692  
Db 601 TGTGGA 606

RESULT 7  
US-09-759-143-206/c  
Sequence 206, Application US/09759143  
Patent No. US2002002248A1

GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Devin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Jiang, Yuqi  
APPLICANT: Henderson, Robert A.  
APPLICANT: Kalos, Michael D.  
APPLICANT: Fanger, Gary R.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.  
APPLICANT: Day, Craig H.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Hepler, William  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.427C23  
CURRENT APPLICATION NUMBER: US/09/759,143  
CURRENT FILING DATE: 2001-01-12  
NUMBER OF SEQ ID NOS: 934  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 206  
LENGTH: 487  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(487)  
OTHER INFORMATION: n = A,T,C or G  
US-09-759-143-206

Query Match 37.4%; Score 439.6; DB 9; Length 487;  
Best Local Similarity 96.0%; Pred. No. 1.7e-88;  
Matches 453; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 536 TTTTGAACCAAGTGCACAAATGCCACAAATGGCAGAGCCCTTGAGGAACGGTTCG 595  
Db 487 TTTTGAACCAAGTGCACAAATGGCAGAGCCCTTGAGGAACGGTTCG 428

QY 596 AAGAGTTCCTTACCGAGGATAGGTGAGATCATTTGATTCAGACAGACACAGTCAATCT 655  
Db 427 AAGAGTTCCTTACCGAGGATAGGTGAGATCATTTGATTCAGACAGACACAGTCAATCT 368

QY 656 TCACCGAAGCCCAAGCCCTAGC-TCATCTGCTGTTGATTTAGATTGTTGATGCAATTC 714  
Db 367 TCACCGAAGCCCAAGCCCTAGCTTCATCTGCTGTTGATTTGATGCAATTC 308

QY 715 TAACCAACTCACACATATACAAAAATCAACATGGGGATGGAGAAGAAATAGCGTTTG 774  
Db 307 TNACCAACTCACACATATACAAATATCNACATGGGGATGGAGAAGAAATAGCGTTTG 248

QY 775 CAGAGTGTATCATCTACTTAATAAATTAACATAATGTTGCTGCTTCATAGTTGGTGG 834  
Db 247 CAGAGTGTATNATCTACTTAATAAATTAACATAATGTTGCTGCTTCATAGTTGGTGG 188

QY 835 AGAAGGACACATCCACTCTTGGAGGAATATTTTACTCAATAATGGCACCTTACATTTA 894  
Db 187 AGAAGGACACATCCACTCTTGGAGGAATATNTACTCAATAATGGCACCTTACATTTA 128

QY 895 TAAATTTGAACAGTCTCTTAATAAGCTTTCTTTTAAATTTAAATATATGTAAGTTTCAGAGCTA 954  
Db 127 TAAATTTGACAGTCTCTTAATAAGCTTTCTTTTAAATTTAAATATATGTAAGTTTCAGAGCTA 68

QY 955 ATAAATGAATGACCAAGACTTTTAAATATAATAAATAAATAAAGAACTTGACTA 1006  
Db 67 ATAAATGAATGACCAAGACTTTTAAATATAATAAATAAATAAAGAACTTGACTA 16

RESULT 8

US-09-780-669-206/c  
 ; Sequence 206, Application US/09780669  
 ; Patent No. US2002005197A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Jiang, Yuqi  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Kalos, Michael D.  
 ; APPLICANT: Fanger, Gary R.  
 ; APPLICANT: Retter, Marc W.  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Day, Craig H.  
 ; APPLICANT: Vedvick, Thomas S.  
 ; APPLICANT: Carter, Darrick  
 ; APPLICANT: Li, Samuel  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Skeiky, Yasir A.W.  
 ; APPLICANT: Hepler, William  
 ; APPLICANT: Hural, John  
 ; APPLICANT: McNeill, Patricia D.  
 ; APPLICANT: Houghton, Raymond L.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
 ; FILE REFERENCES: 210121.427C24  
 ; CURRENT APPLICATION NUMBER: US/09/780,669  
 ; CURRENT FILING DATE: 2001-02-09  
 ; NUMBER OF SEQ ID NOS: 943  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 206  
 ; LENGTH: 487  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: (1)...(487)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-780-669-206

Query Match 37.4%; Score 439.6; DB 9; Length 487;  
 Best Local Similarity 96.0%; Pred. No. 1.7e-88;  
 Matches 453; Conservative 0; Mismatches 18; Indels 1; Gaps 1;  
 QY 536 TTTTGAACAAGTGCAAGATGCCCAAAATGTCAGAGCGCTTTGAGGAAGCGTTTCG 595  
 Db 487 TTTTGAACAAGTGCAAGATGCCCAAAATGTCAGAGCGCTTTGAGGAAGCGTTTCG 428  
 QY 596 AAGAGTTCTTGTCTACCGAGGATAGGTTCAGATCATTTGATTCAGACAGACACAGTCAATCT 655  
 Db 427 AAGAGTTCTTGTCTACCGAGGATAGGTTCAGATCATTTGATTCAGACAGACACAGTCAATCT 368  
 QY 656 TCACCGAAGCCCAAGCCCTAGC-TCATCTTGTCTGTGATTTAGATTGTTGATGCATTC 714  
 Db 367 TCACCGAAGCCCAAGCCCTAGC-TCATCTTGTCTGTGATTTAGATTGTTGATGCATTC 308  
 QY 715 TAACCAACTCACAATATACAAAATCAACATGCGGATGGAGAGAAATAGCGTTTG 774  
 Db 307 TNACCAACTCACAATATACAAAATCAACATGCGGATGGAGAGAAATAGCGTTTG 248  
 QY 775 CAGCAGTGATCATCTACTTAATAAATTAACCTAATGTTGCTGCTTCAATTAGTTGGTGG 834  
 Db 247 CAGCAGTGATCATCTACTTAATAAATTAACCTAATGTTGCTGCTTCAATTAGTTGGTGG 188  
 QY 835 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATATGGCACCTTACATTTA 894  
 Db 187 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATATGGCACCTTACATTTA 128  
 QY 895 TAAATTTGAACAGTTGTCTTAATAAGCTTTCTTTAATTTAAATATGTAAGTTGCAGAGCTA 954  
 Db 127 TAAATTTGAACAGTTGTCTTAATAAGCTTTCTTTAATTTAAATATGTAAGTTGCAGAGCTA 68

QY 955 ATAAATGAATGACCAAGACTTTTAAATATATATAAATAAAGAACTTGACTA 1006  
 Db 67 ATAAATGAATGACCAAGACTTTTAAATATATATAAATAAAGAACTTGACTA 16  
 RESULT 9  
 US-09-030-606-206/c  
 ; Sequence 206, Application US/09030606  
 ; Patent No. US20020081580A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF PROSTATE CANCER AND METHODS  
 ; NUMBER OF SEQUENCES: 224  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SEED and BERRY LLP  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: WA  
 ; COUNTRY: USA  
 ; ZIP: 98104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/030,606  
 ; FILING DATE: 25-FEB-1998  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Maki, David J.  
 ; REGISTRATION NUMBER: 31,392  
 ; REFERENCE/DOCKET NUMBER: 210121.428C3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 206:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 487 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 US-09-030-606-206

Query Match 37.4%; Score 439.6; DB 9; Length 487;  
 Best Local Similarity 96.0%; Pred. No. 1.7e-88;  
 Matches 453; Conservative 0; Mismatches 18; Indels 1; Gaps 1;  
 QY 536 TTTTGAACAAGTGCAAGATGCCCAAAATGTCAGAGCGCTTTGAGGAAGCGTTTCG 595  
 Db 487 TTTTGAACAAGTGCAAGATGCCCAAAATGTCAGAGCGCTTTGAGGAAGCGTTTCG 428  
 QY 596 AAGAGTTCTTGTCTACCGAGGATAGGTTCAGATCATTTGATTCAGACAGACACAGTCAATCT 655  
 Db 427 AAGAGTTCTTGTCTACCGAGGATAGGTTCAGATCATTTGATTCAGACAGACACAGTCAATCT 368  
 QY 656 TCACCGAAGCCCAAGCCCTAGC-TCATCTTGTCTGTGATTTAGATTGTTGATGCATTC 714  
 Db 367 TCACCGAAGCCCAAGCCCTAGC-TCATCTTGTCTGTGATTTAGATTGTTGATGCATTC 308  
 QY 715 TAACCAACTCACAATATACAAAATCAACATGCGGATGGAGAGAAATAGCGTTTG 774  
 Db 307 TNACCAACTCACAATATACAAAATCAACATGCGGATGGAGAGAAATAGCGTTTG 248  
 QY 775 CAGCAGTGATCATCTACTTAATAAATTAACCTAATGTTGCTGCTTCAATTAGTTGGTGG 834  
 Db 247 CAGCAGTGATCATCTACTTAATAAATTAACCTAATGTTGCTGCTTCAATTAGTTGGTGG 188  
 QY 835 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATATGGCACCTTACATTTA 894  
 Db 187 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATATGGCACCTTACATTTA 128





```
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 206
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(487)
; OTHER INFORMATION: n = A,T,C or G
US-09-895-814-206

Query Match      37.4%; Score 439.6; DB 10; Length 487;
Best Local Similarity 96.0%; Pred. No. 1.7e-88;
Matches 453; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 536 TTTTGAACAAGTGCAAAAGATGCCAAATGTGGCAGAGCGCTTTGAGGAAGCGGTTGC 595
DB 487 TTTTGAACAAGTGCAAAAGATGCCAAATGTGGCAGAGCGCTTTGAGGAAGCGGTTGC 428
QY 596 AAGAGTTCTTGTACCGAGGATAGGTGAGTTCAGATCATTTGATTCAGACAGACAGTCAATCT 655
DB 427 AAGAGTTCTTGTACCGAGGATAGGTGAGTTCAGATCATTTGATTCAGACAGACAGTCTATTT 368
QY 715 TAACCAACTCACACATATACAAAATCAACATGGGATGGAGAGAGAAATAGCGTTTG 774
DB 307 TNACCAACTCACACATATACAAAATCAACATGGGATGGAGAGAGAAATAGCGTTTG 248
QY 775 CAGCAGTGTATCATCTACTAATAAATTAATACTAATGTTGCTGCTTCATTAGTTGGTGG 834
DB 247 CAGCAGTGTATCATCTACTAATAAATTAATACTAATGTTGCTGCTTCATTAGTTGGTGG 188
QY 835 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTA 894
DB 187 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTA 128
QY 895 TAAATTTGAACAGTTGCTTAATAACCTTTCTTTTAAATTTAAATATGTAAGTTGCAGAGCTA 954
DB 127 TAAATTTGAACAGTTGCTTAATAACCTTTCTTTTAAATTTAAATATGTAAGTTGCAGAGCTA 68
QY 955 ATAAATGAATGACCAAGCTTTTAAATTAATAAATAAAGAACTTGACTA 1006
DB 67 ATAAATGAATGACCAAGCTTTTAAATTAATAAATAAAGAACTTGACTA 16
```

## RESULT 15

```
US-10-144-678A-206/C
; Sequence 206, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yufu
```

```
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Hepler, William T.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals y de Bassols, Carlota
; APPLICANT: Foy, Teresa M.
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Deng, Ta
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C28
; CURRENT APPLICATION NUMBER: US/10/144,678A
; CURRENT FILING DATE: 2002-08-12
; NUMBER OF SEQ ID NOS: 1033
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 206
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 29, 106, 119, 154, 204, 236, 279, 306, 318, 325, 353, 448,
; LOCATION: 466, 470
; OTHER INFORMATION: n = A,T,C or G
US-10-144-678A-206

Query Match      37.4%; Score 439.6; DB 13; Length 487;
Best Local Similarity 96.0%; Pred. No. 1.7e-88;
Matches 453; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 536 TTTTGAACAAGTGCAAAAGATGCCAAATGTGGCAGAGCGCTTTGAGGAAGCGGTTGC 595
DB 487 TTTTGAACAAGTGCAAAAGATGCCAAATGTGGCAGAGCGCTTTGAGGAAGCGGTTGC 428
QY 596 AAGAGTTCTTGTACCGAGGATAGGTGAGTTCAGATCATTTGATTCAGACAGACAGTCAATCT 655
DB 427 AAGAGTTCTTGTACCGAGGATAGGTGAGTTCAGATCATTTGATTCAGACAGACAGTCTATTT 368
QY 656 TCACCGAAGCCCAAGCCTAGC-TCATCTTGTGTTGATTGTAGATTGTTGATGCTATTC 714
DB 367 TCACCGAAGCCCAAGCCTAGCCTTCATCTTGTGTTGATTGTAGATTGTTGATGCTATTC 308
QY 715 TAACCAACTCACACATATACAAAATCAACATGGGATGGAGAGAGAAATAGCGTTTG 774
DB 307 TNACCAACTCACACATATACAAAATCAACATGGGATGGAGAGAGAAATAGCGTTTG 248
QY 775 CAGCAGTGTATCATCTACTAATAAATTAATACTAATGTTGCTGCTTCATTAGTTGGTGG 834
DB 247 CAGCAGTGTATCATCTACTAATAAATTAATACTAATGTTGCTGCTTCATTAGTTGGTGG 188
QY 835 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTA 894
DB 187 AGAAGGGACACATCCACTCTTGGAGGAATATATTTACTCAATAATGGCACCTTACATTTA 128
QY 895 TAAATTTGAACAGTTGCTTAATAACCTTTCTTTTAAATTTAAATATGTAAGTTGCAGAGCTA 954
DB 127 TAAATTTGAACAGTTGCTTAATAACCTTTCTTTTAAATTTAAATATGTAAGTTGCAGAGCTA 68
QY 955 ATAAATGAATGACCAAGCTTTTAAATTAATAAATAAAGAACTTGACTA 1006
DB 67 ATAAATGAATGACCAAGCTTTTAAATTAATAAATAAAGAACTTGACTA 16
```



Thu Feb 19 09:29:46 2004

us-09-988-974-6.rnpb

Page 11

Search completed: February 10, 2004, 04:54:25  
Job time : 494 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2004, 00:45:41 ; Search time 2885 Seconds

(without alignments)  
9898.703 Million cell updates/sec

Title: US-09-988-974-6

Perfect score: 1175

Sequence: 1 CACTGTGATGAAACACTTTT.....GTAAAAA.....1175

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estm.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_estc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_est3.\*

12: gb\_est4.\*

13: gb\_est5.\*

14: gb\_est6.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pin.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_pbg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	888	75.6	1201	13	BX422913
2	832.4	70.8	1201	13	BX422914
3	808.2	68.8	1201	13	BX361467
4	802.2	68.3	837	13	BQ221697

5	790.2	67.3	835	12	BI761520
6	720.4	61.3	791	9	AI671547
C 7	713.2	60.7	781	9	AI934576
C 8	712	60.6	734	9	AI833027
C 9	707.6	60.2	728	12	BM981745
C 10	706.6	60.1	1027	13	BX344135
C 11	704.4	59.9	720	13	BQ772388
C 12	704	59.9	881	14	CD515129
C 13	692.2	58.9	767	9	AI479232
C 14	688.6	58.6	778	9	AW007521
C 15	686	58.4	708	9	AI758228
C 16	685.6	58.3	714	14	CA308701
C 17	679.8	57.9	696	9	AI659875
C 18	675	57.4	715	9	AI763029
C 19	667.4	56.8	705	10	EG054716
C 20	664.8	56.6	734	9	AI765395
C 21	659.4	56.1	678	13	BU682548
C 22	659.2	56.1	1205	11	AK032133
C 23	659.2	56.1	1242	11	AK017301
C 24	659	56.1	659	14	CA441619
C 25	653.8	55.6	1072	12	BM476930
C 26	635.6	54.1	682	10	BE326331
C 27	617.6	52.6	823	14	CA488998
C 28	617	52.5	658	10	BE858526
C 29	616.4	52.5	910	12	BI825651
C 30	609	51.8	609	12	BM353471
C 31	605.6	51.5	613	9	AI922378
C 32	604.4	51.4	656	9	AI423136
C 33	593	50.5	887	13	BX352178
C 34	592	50.4	720	12	BI818653
C 35	589.4	50.2	614	12	BM674469
C 36	588	50.0	609	12	BM717088
C 37	585	49.8	615	13	BU616781
C 38	583.4	49.7	615	12	BG874452
C 39	582.8	49.6	1179	12	BM926730
C 40	575.4	49.0	588	9	AW614366
C 41	566.8	48.2	576	12	BM126004
C 42	566.6	48.2	761	12	BI766528
C 43	565.2	48.1	571	9	AI281626
C 44	561	47.7	561	9	AA563671
C 45	560.8	47.7	573	10	BE464245

#### ALIGNMENTS

RESULT 1  
BX422913/C  
LOCUS BX422913 Homo sapiens ADULT BRAIN Homo sapiens cDNA clone  
DEFINITION CSODN002YB16 3-PRIME, mRNA sequence.  
ACCESSION BX422913  
VERSION BX422913.1 GI:30655413  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Li, W.B., Gruber, C., Jesses, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 11038.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cg1-bin/cluster.cgi?seq=CSODN002DA08NP1&cluster=11038.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600



Db 279 AGTTTGATACCCAGCTCTTCCATCAATAGGTGGAAATTTTAAATAAARATTTGGAAG 338  
 Qy 238 TGGATGACATTTTGTACCATGAGATTTGGGACACGGCAGGTTCAGGAGCGATTCGAA 297  
 Db 339 TGGATGACATTTTGTACCATGAGATTTGGGACACGGCAGGTTCAGGAGCGATTCGAA 398  
 Qy 298 GCCTGAGGACACCATTTTACAGAGGTTCTGACTGCTGCTGCTGCTTACCTTTAGTGTGATG 357  
 Db 399 GCCTGAGGACACCATTTTACAGAGGTTCTGACTGCTGCTGCTTACCTTTAGTGTGATG 458  
 Qy 358 ATTACAAAGCTTCCAGACCTTAAGTAACCTGGAGAAAGAAATTCATATATTATGCGAGATG 417  
 Db 459 ATTACAAAGCTTCCAGACCTTAAGTAACCTGGAGAAAGAAATTCATATATTATGCGAGATG 518  
 Qy 418 TGAAGAGCCTGAGAGCTTCTCTTTCTGATCTCTGGGTACCAAGATTGACATAAGCGAAC 477  
 Db 519 TGAAGAGCCTGAGAGCTTCTCTTTCTGATCTCTGGGTACCAAGATTGACATAAGCGAAC 578  
 Qy 478 GGCAGGTGTCTACAGAAAGCCCAAGCTTGGTGCAGGGCAACACGGCGACTATCCTTATT 537  
 Db 579 GGCAGGTGTCTACAGAAAGCCCAAGCTTGGTGCAGGGCAACACGGCGACTATCCTTATT 638  
 Qy 538 TTGAACAAGTGCAGAAAGATGCCACAAATGTGGCAGAGCCTTTGAGGAAGCGGTTGAA 597  
 Db 639 TTGAACAAGTGCAGAAAGATGCCACAAATGTGGCAGAGCCTTTGAGGAAGCGGTTGAA 698  
 Qy 598 GAGTCTTGTCTACCGAGGATAGTTCAGATCATTTTGAATTCAGACAGACAGATCAATCTTC 657  
 Db 699 GAGTCTTGTCTACCGAGGATAGTTCAGATCATTTTGAATTCAGACAGACAGATCAATCTTC 758  
 Qy 658 ACCGAAGCCCAAGCCTAGCTCATCTGCTGTTGATTTAGATTGTTGATGATTCATTA 717  
 Db 759 ACCGAAGCCCAAGCCTAGCTCATCTGCTGTTGATTTAGATTGTTGATGATTCATTA 818  
 Qy 718 CCAACTCACATATACAAAAATCAACAT-GGGGATGGAGAGAGAAATTAGCGTTTGA 776  
 Db 819 CCAACTCACATATACAAAAATCAACATGGGGATGGAGAGAGAAATTAGCGTTTGA 878  
 Qy 777 GCAGTGTATCATCTATAATAAATAAATAAATGTTGCTTCAATAGTTGGTGGGAG 836  
 Db 879 GCAGTGTATCATCTATAATAAATAAATAAATGTTGCTTCAATAGTTGGTGGGAG 938  
 Qy 837 AAGGGACACATCCACTCTGGAGGATATATTTACTCAATATGGCAGCTTACATTTATA 896  
 Db 939 AAGGGACACATCCACTCTGGAGGATATATTTACTCAATATGGCA-CTTACATTTATA 997  
 Qy 897 AATTGTAACAGTTGTCTAATAACGTTTCTTTAATTTAAATATGTAAGTTGCAGAGCTAAT 956  
 Db 998 AATTGT-AMAGTTGTATAAAGTTCTTTAATTTAAATATGTAAGTTGCAGAGCTAAT 1056  
 Qy 957 AATTGAATGACCAAGACTTTAATTTAATAATAAATAAGAAA 998  
 Db 1057 GAATGACCAAGCTTTAATTTAATAAATAAAGAACTGGAATAA 1098

RESULT 3  
 BX361467  
 LOCUS  
 DEFINITION  
 BX361467 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 clone CSODI085YP05 5-PRIME, mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1201)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France  
 Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 11038.f For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CSODI085CH03QPl&cluster=11038.f. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CSODI085CH03QPl.  
 Location/Qualifiers  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clones="CSODI085YP05"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with NotI and cloned into the NotI and EcoRV  
 sites of the pCMVSPORT 6 vector. Library was normalized."  
 BASE COUNT 347 a 212 c 258 g 326 t 58 others  
 ORIGIN

FEATURES  
source

Query Match 68.8%; Score 808.2; DB 13; Length 1201;  
 Best Local Similarity 92.8%; Pred. No. 1.7e-100;  
 Matches 851; Conservative 24; Mismatches 39; Indels 3; Gaps 3;  
 Qy 1 CACTGTGATGAACACACTTTTCCCGTGTCTGTTGAGTGCATCTTCTCAACACCCCTAGGAG 60  
 Db 121 CACTGTGATGAACACACTTTTCCCGTGTCTGTTGAGTGCATCTTCTCAACACCCCTAGGAG 180  
 Qy 61 GGTCTTGAAGCTTTTTCAGATTAACTGCGAGGAAATCATCTTTTAAAGTAATTC 120  
 Db 181 GGTCTTGAAGCTTTTTCAGATTAACTGCGAGGAAATCATCTTTTAAAGTAATTC 240  
 Qy 121 TCCTTGGAGATGTTGGAGTTGGGAGAGTTTCACTTATGAACAGATATGTAACATAAAGT 180  
 Db 241 TCCTTGGAGATGTTGGAGTTGGGAGAGTTTCACTTATGAACAGATATGTAACATAAAGT 300  
 Qy 181 TTGATACCCAGCTTCTCCATACAAATAGGTGTGGAATTTTAAATAAAGATTTGGAGTGG 240  
 Db 301 YTGATACCCAGCTTCTCCATACAAATAGGTGTGGAATTTTAAATAAAGATTTGGAGTGG 360  
 Qy 241 ATGACACATTTGTTTACCATGCAGATTTGGGACACGGCAGGTTCAGGCGGATTCGAGGCC 300  
 Db 361 ATGACACATTTGTTTACCATGCAGATTTGGGACACGGCAGGTTCAGGCGGATTCGAGGCC 420  
 Qy 301 TGAGGACACCATTTTACAGAGGTTCTGACTGCTGCTTACTTTTGTGTCGATGATT 360  
 Db 421 TGAGGACACCATTTTACAGAGGTTCTGACTGCTGCTTACTTTTGTGTCGATGATT 480  
 Qy 361 CACAAAGCTTCCAGAACTTAAGTAAGTGAAGAAAGAAATTCATATATATGTCGATGTA 420  
 Db 481 CACAAAGCTTCCAGAACTTAAGTAAGTGAAGAAAGAAATTCATATATATGTCGATGTA 540  
 Qy 421 AAGAGCCTGAGAGCTTTTCCCTTTTGTGATTTCTGGGTAACTGAGATTTGACATAAGCGACGC 480  
 Db 541 AAGAGCCTGAGAGCTTTTCCCTTTTGTGATTTCTGGGTAACTGAGATTTGACATAAGCGACGC 600  
 Qy 481 AGTGTCTACAGAAAGACCCAAAGCTTTGGTGCAGGACAAACGGCGACTATTCCTTATTTTG 540  
 Db 601 AGTGTCTACAGAAAGACCCAAAGCTTTGGTGCAGGACAAACGGCGACTATTCCTTATTTTG 660  
 Qy 541 AACAAGTCAAAAGATGCCACAAATGTGGCAGACGCC-TTTGAGGAAGCGGTTCCGAAGA 599  
 Db 661 AACAAGTCAAAAGATGCCACAAATGTGGCAGACGCCCTTTGAGGAAGCGGTTCCGAAGA 720  
 Qy 600 GTTCTTGTCTACCGAGGATAGGTTCAGATCATTTGATTCAGACAGACAGTCAATCTTCAC 659  
 Db 721 GTTCTTGTCTACCGAGGATAGGTTCAGATCATTTGATTCAGACAGACAGTCAATCTTCAC 780



DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LRAM1466 row: n column: 15  
 High quality sequence stop: 835.

FEATURES  
 Location/Qualifiers  
 1..835

source  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clones="IMAGE:5186894"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH MGC 116"  
 /note="Organ: pooled colon, kidney, stomach; Vector:  
 PCMV-SPORE6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
 source anonymous pool of 3 colons, age 26 yo male, 49 yo  
 female, 71 yo male colon; 46 yo male kidney, and pool of 2  
 stomachs, 62 yo male and 70 yo female. Library is  
 oligo-dT primed and directionally cloned (EcoRV site is  
 destroyed upon cloning). Average insert size 1.4 kb,  
 insert size range 1-3 kb. Library is normalized and  
 enriched for full-length clones and was constructed by C.  
 Gruber (Invitrogen). Research Genetics tracking code  
 023. Note: this is a NIH MGC Library."  
 BASE COUNT 269 a 145 c 186 g 235 t  
 ORIGIN

Query Match 67.3%; Score 790.2; DB 12; Length 835;  
 Best Local Similarity 98.8%; Pred. No. 5.5e-98;  
 Matches 81; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

231 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 290  
 2 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 60  
 291 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 350  
 61 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 120  
 351 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 410  
 121 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 180  
 411 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 470  
 181 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 240  
 471 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 530  
 241 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 300  
 531 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 590  
 301 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 360  
 591 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 650  
 361 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 710  
 651 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 770  
 421 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 480  
 771 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 830  
 481 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 540  
 831 TTGGAAGTGGATGACATTTTGTACCATCGATTGGGACACGGCAGGTGAGGACGA 890

Db 601 TGGGAGAGGGACACATCCACTCTGGAGGAATATATTTACTCAATAATGGCACCTTACA 660  
 QY 891 TTTATAAATTTGACAGTTGCTATAAAGCTTTCTTTAATTTAAATATGTAAGTTGCAGA 950  
 Db 661 TTTATAAATTTGACAGTTGCTATAAAGCTTTCTTTAATTTAAATATGTAAGTTGCAGA 720  
 QY 951 GCTAATAAATGAATGACCAAGACTTTAA-TTATAATAAAAATAGAAACTTGACTATTC 1009  
 Db 721 GCTAATAAATGAATGACCAAGACTTTAAAGTCATAATAAATAAGAACTTGACTATTC 780  
 QY 1010 TAGAAGTTATATCTGCTGATTTTCTCTGGGAAAATGGGAACTACTTTT 1056  
 Db 781 TAGAAGTTATATCTGCTGATTTTCTCTGGGAAAATGGGAACTACTTTT 827

## RESULT 6

AI671547/c

LOCUS

DEFINITION

AI671547

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

## FEATURES

source

1..791

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clones="IMAGE:2307492"

/tissue\_type="pooled germ cell tumors"

/lab\_host="DH10B"

/clone\_lib="NCI CGAP GC6"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA

from the normalized library NCI-CGAP GC4 was prepared, and

ss circles were made in vitro. Following HAP purification, and

this DNA was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from a pool

of 5,000 clones made from the same library (clones IDs

1257096-1258631, 1469064-1470983, and 1475592-1476743).

Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 214 a 170 c 139 g 267 t

ORIGIN

Query Match 61.3%; Score 720.4; DB 9; Length 791;

Best Local Similarity 98.1%; Pred. No. 1.6e-88;

Matches 760; Conservative 0; Mismatches 12; Indels 3; Gaps 3;

229 ATTTGGAGTGGATGGATTTTGTACCATGAGATTTGGGACACGGCAGGTCCAGAGC 288  
 Db |||||  
 790 ATTTGGAGTGGATGGACATTTTGTACCATGAGATTTGGGACACGGCAGGTCCAGAGC 733  
 Qy |||||  
 289 GATTCCGAAGCCTGAGGACACCAATTTTACAGAGTCTGACTGCTGCTTACCTTTTA 348  
 Db |||||  
 732 GATCCGAAGCCTGAGGACACATTTTACAGAGTCTGACTGCTGCTTACCTTTTA 674  
 Qy |||||  
 349 GTGTGATGATTCACAAAGCTTCCAGACTTAAGTAACTGGAAGAAGAAATTCATATTT 408  
 Db |||||  
 673 GTGTGATGATTCACAAAGCTTCCAGACTTAAGTAACTGGAAGAAGAAATTCATATTT 614  
 Qy |||||  
 409 ATGCAGATGTGAAGAGCCTGAGAGCTTCTCTTTGTGATTCCTGGGTAAACAAGATTGACA 468  
 Db |||||  
 613 ATGCAGATGTGAAGAGCCTGAGAGCTTCTCTTTGTGATTCCTGGGTAAACAAGATTGACA 554  
 Qy |||||  
 469 TAAGCGAAGCGAGGTGTCTACAGAGAAGCCCAAGCTTGGTGCAGGGAACAGCGGACT 528  
 Db |||||  
 553 TAAGCGAAGCGAGGTGTCTACAGAGAAGCCCAAGCTTGGTGCAGGGAACAGCGGACT 494  
 Qy |||||  
 529 ATCTTATTTTGAACAAGTGCACAAAGATGCCCAAAATGTGGCAGCAGCTTTGGGAAG 588  
 Db |||||  
 493 ATCTTATTTTGAACAAGTGCACAAAGATGCCCAAAATGTGGCAGCAGCTTTGGGAAG 434  
 Qy |||||  
 589 CGGTTCGAAGAGTCTTGTCTACCGAGGATAGGTGAGATCATTTGATTCAGACACACAG 648  
 Db |||||  
 433 CGGTTCGAAGAGTCTTGTCTACCGAGGATAGGTGAGATCATTTGATTCAGACACACAG 374  
 Qy |||||  
 649 TCATCTTCCAGGAAGCCCAAGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 708  
 Db |||||  
 373 TCATCTTCCAGGAAGCCCAAGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 314  
 Qy |||||  
 709 GCATTTCAACAACTCACACATATACACAAATCAACATGGGATGGAGAGAGAAATTAG 768  
 Db |||||  
 313 GCATTTCAACAACTCACACATATACACAAATCAACATGGGATGGAGAGAGAAATTAG 254  
 Qy |||||  
 769 CGTTGCGAGCAGTGTATCTACTACTATAAATAAATAAATAAATAAATAAATAAATAAATA 828  
 Db |||||  
 253 CGTTGCGAGCAGTGTATCTACTACTATAAATAAATAAATAAATAAATAAATAAATAAATA 194  
 Qy |||||  
 829 GGTGGGAGAGGACACATCCATCTTGGAGGATATATTTACTCAATATATGACACCTTA 888  
 Db |||||  
 193 GGTGGGAGAGGACACATCCATCTTGGAGGATATATTTACTCAATATATGACACCTTA 134  
 Qy |||||  
 889 CATTATTAATGTAACAGTGTCTTAATACAGTTCCTTTAAATTAATTAATTAATTAATTAAT 948  
 Db |||||  
 133 CATTATTAATGTAACAGTGTCTTAATACAGTTCCTTTAAATTAATTAATTAATTAATTAAT 74  
 Qy |||||  
 949 GAGCTAATAATGAATGACCAAGACTTTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1003  
 Db |||||  
 73 GAGCTAATAATGAATGACCAAGACTTTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19

RESULT 7  
 AI934576/c  
 LOCUS  
 DEFINITION  
 wp09f06.x1 NCI CGAP Kid12 Homo sapiens cDNA clone IMAGE:2464355 3,  
 similar to SW:RAB9\_HUMAN P5151 RAS-RELATED PROTEIN RAB-9.; mRNA  
 sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 781)  
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapps-1@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.lnl.gov/bbrp/image/image.html  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 469.  
 Location/Qualifiers  
 1. 781  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2464355"  
 /tissue\_type="2 pooled tumors (clear cell type)"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Kid12"  
 /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
 a modified polylinker; Site 1: Not 1; Site 2: Eco RI;  
 Plasmid DNA from the normalized library NCI-CCAP Kids was  
 prepared, and ss circles were made in vitro. Following HAP  
 purification, this DNA was used as tracer in a subtractive  
 hybridization reaction. The driver was PCR-amplified cDNAs  
 from a pool of 5,000 clones made from the same library  
 (cloneID8 1323912-1325831, 1471368-1472903 and  
 1492104-1493255). Subtraction by Bento Soares and M.  
 Fatima Bonaldo."

BASE COUNT 235 a 155 c 122 g 263 t 6 others  
 ORIGIN  
 Query Match 60.7%; Score 713.2; DB 9; Length 781;  
 Best Local Similarity 96.7%; Pred. No. 1.5e-87;  
 Matches 757; Conservative 0; Mismatches 23; Indels 3; Gaps 3;  
 Qy 390 AAGAAAGATTTCATATATATGACAGATGTGAAAGAGCCTGAGAGCTTCCCTTTGTGATT 449  
 Db 781 AAGAAAGATTTCATATATATGACAGATGTGAAAGAGCCTGAGAGCTTCCCTTTGTGATT 722  
 Qy 450 CTGGTTACAGAGTTGACATAGCGACGCGAGTGTCTACAGAG-AAGCCCAAGCTTG 508  
 Db 721 CTGGGT-ACAGATTGACATAGCGACGCGAGTGTCTACAGAGNAAGCCCAAGCTTG 663  
 Qy 509 GTGCAGGACAAACCGGACCTATCCTTATTTTGAACAGTGCACCAAGATGCCCAAGTGT 568  
 Db 662 GTGCAGGACAAACCGGACCTATCCTTATTTG-ACCAAGTGCACCAAGATGCCCAAGTGT 604  
 Qy 569 GGCAGCAGCCTTTGAGGAAGGGTTCGAAGAGTCTTGTCTACCGAGGATAGGTGAGTCA 628  
 Db 603 GGCAGCAGCCTTTGAGGAAGGGTTCGAAGAGTCTTGTCTACCGAGGATAGGTGAGTCA 544  
 Qy 629 TTTGATTTCAGACACACAGTCAATCTTCACGAAGCCCAAGCCTAGCTCATCTTGTCTG 688  
 Db 543 TTTGATTTCAGACACACAGTCAATCTTCACGAAGCCCAAGCCTAGCTCATCTTGTCTG 484  
 Qy 589 TTGATTGTTAGATTGTTGATGCTTTCACCAACTCACAATATACACAAAATCAACATG 748  
 Db 483 TTGATTGTTAGATTGTTGATGCTTTCACCAACTCACAATATACACAAAATCAACATG 424  
 Qy 749 GGGATGGAAGAAGAAATTAGCGTTTTCGAGCAGTGTATCATCTACTAATAAATAAATAA 808  
 Db 423 GGGATGGAAGAAGAAATTAGCGTTTTCGAGCAGTGTATCATCTACTAATAAATAAATAA 364  
 Qy 809 ATGTTGCTGCTTCATTAGTTGGTGGGAGAGGAGGACATCCACTCTTTGGAGGAATATAT 868  
 Db 363 ATGTTGCTGCTTCATTAGTTGGTGGGAGAGGAGGACATCCACTCTTTGGAGGAATATAT 304  
 Qy 869 TACTCAATATGACACCTTACATTTAATTTGACAGTGTCTAATAACGTTTCTTTTA 928  
 Db 303 TACTCAATATGACACCTTACATTTAATTTAATAATTTGAACGTTGTCTAATAACGTTTCTTT 244  
 Qy 929 ATTTAATAATGTAAGTTGACAGCTAATAAATGAATGACCAAGACTTTTAATAATAA 988

FEATURES  
 source





JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548  
Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library prepared by: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.regen.com) or from Open Biosystems  
(www.openbiosystems.com).  
The following repetitive elements were found in this cDNA  
sequence: 1-48, >AT rich#Low\_complexity (matched complement)  
Seq primer: M13 FORWARD  
POLYA=yes

FEATURES  
source

1. .728  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-EN1-adi-p-23-0-UI"  
/tissue\_type="Primary Lung Cystic Fibrosis Epithelial  
Cells"  
/dev\_stage="Adult"  
/lab\_host="PH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-EN1"  
/note="Organ: Lung; Vector: pVT3-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-CF-EN1 is a normalized cDNA library containing the  
following tissue(s): Primary Lung Cystic Fibrosis  
Epithelial Cells. The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was ligated to an EcoR I adaptor, digested with Not  
I, and cloned directionally into pVT3-Pac vector. The  
oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (GT)18 tail. The  
sequence tag for this library is CTGCTCAGGT.  
TAG LIB=UI-CF-EN1  
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS  
6hr to LPS 24h  
TAG\_SEQ=CTGCTCAGGT"

BASE COUNT 220 a 140 c 111 g 257 t  
ORIGIN

Query Match 60.2%; Score 707.6; DB 12; Length 728;  
Best Local Similarity 98.8%; Pred. No. 9.1e-87;  
Matches 713; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 454 GTAACAAGATTGACATAAGCGAGCGAGTGTCTACAGAGAAGCCCAAGCTTGTCGA 513  
DB 728 GTAACAAGATTGACATAAGCGAGCGAGTGTCTACAGAGAAGCCCAAGCTTGTCGA 669  
QY 514 GGGACAAACCGGACATCTCTTTTGAACAAGTGCAGAGATGCCCAATGTGGCAG 573  
DB 668 GGGACAAACCGGACATCTCTTTTGAACAAGTGCAGAGATGCCCAATGTGGCAG 609  
QY 574 CAGCCTTTGAGGAAGCGGTTCGAAGAGTTCTTGTACCGAGGATAGGTTCAGATCATTTGA 633  
DB 608 CAGCCTTTGAGGAAGCGGTTCGAAGAGTTCTTGTACCGAGGATAGGTTCAGATCATTTGA 549  
QY 634 TTCAGACACACAGTCATCTTCCAGCAAGCCCAAGCTAGCTCATCTTCTGCTTGAT 693  
DB 548 TTCAGACACACAGTCATCTTCCAGCAAGCCCAAGCTAGCTCATCTTCTGCTTGAT 489

QY 694 TGTTAGATTCTTGATGCATTTCTAACCAACTCACAATATACACAAAATCAACATGGGGAT 753  
DB 488 TGTTAGATTCTTGATGCATTTCTAACCAACTCACAATATACACAAAATCAACATGGGGAT 429  
QY 754 GGGAGAGAAATTAGCGTTTTCAGAGAGTGTATCTACTACTATAAATAAATAAATAATGTT 813  
DB 428 GGGAGAGAAATTAGCGTTTTCAGAGAGTGTATCTACTACTATAAATAAATAAATAATGTT 369  
QY 814 GCTGCTTCATTAGTTGGTGGGAGAGGGACACATCCACTCTTTCGAGGAATATATTTTACTC 873  
DB 368 GCTGCTTCATTAGTTGGTGGGAGAGGGACACATCCACTCTTTCGAGGAATATATTTTACTC 309  
QY 874 AATAATGGCACCTTACATTTTATAAATTGTAACAGTTGTCTATAAGCTTCTTTAATTTA 933  
DB 308 AATAATGGCACCTTACATTTTATAAATTGTAACAGTTGTCTATAAGCTTCTTTAATTTA 249  
QY 934 AATATGTAAGTTGCAGAGCTTAATAAATGAAATGACCAAGACTTTTAATTATAAATAAATA 993  
DB 248 AATATGTAAGTTGCAGAGCTTAATAAATGAAATGACCAAGACTTTTAATTATAAATAAATA 189  
QY 994 AGAACTTGACTATTTTAGAAGTTATACCTGGATTTTTCCTGGGAAATGGAGAACTAC 1053  
DB 188 AGAACTTGACTATTTTAGAAGTTATACCTGGATTTTTCCTGGGAAATGGAGAACTAC 129  
QY 1054 TTTTATATGTTGATGTTTTTATGCAATTAGCAATTCGATTCTTGGTTCAGGGAATACCT 1113  
DB 128 TTTTATATGTTGATGTTTTTATGCAATTAGCAATTCGATTCTTGGTTCAGGGAATACCT 69  
QY 1114 TCCTAAAGCAATTAATGTTAGATATTAAGATTAAATCTAATGTTAAATAAATAAATAAATA 1173  
DB 68 TCCTAAAGCAATTAATGTTAGATATTAAGATTAAATCTAATGTTAAATAAATAAATAAATA 9  
QY 1174 AA 1175  
DB 8 AA 7  
RESULT 10  
BX344135  
LOCUS  
DEFINITION BX344135 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
Clone CS0D1060YA02 5-PRIME, mRNA sequence.  
ACCESSION BX344135  
VERSION BX344135.1 GI:30338244  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1027)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 11038.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS1A1015ZG02QP1&cluster=11038.f. Contact :  
Peng Liang Email : filiang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS1A1015ZG02QP1.  
FEATURES  
1. .1027  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1060YA02"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT	267 a	222 c	193 g	289 t	56 others
ORIGIN					
Query Match	60.1%; Score 706.6; DB 13; Length 1027;				
Best Local Similarity	89.1%; Pred. No. 1e-86;				
Matches	698; Conservative 50; Mismatches 35; Indels 0; Gaps 0;				
QY	1	CACGTGATGAACACATTTTCCCGTTCGTTTGAAGTGCATCTTCTCAACAACCCCTAGGAG	60		
DB	245	CACCTTTATKAACACATTTTCCCGTTCGTTTGAAGTGCATCTTCTCAACAACCCCTAGGAG	304		
QY	61	GGTCTTGAAGCTTTTGAATTAACATGCGAGGAAATCATCATTTTAAAGTAATTC	120		
DB	305	GGTCTTGAAGCTTTTGAATTAACATGCGAGGAAATCATCATTTTAAAGTAATTC	364		
QY	121	TCCTTGGAGATGGTGGAGTGGGAGGTTCACTTATGAACAGATATGTAACATAAGT	180		
DB	365	TCCTTGGAGATGGTGGAGTGGGAGGTTCACTTATGAACAGATATGTAACATAAGT	424		
QY	181	TTGATACCCAGCTCTCCATACAAATAGGTGCGAATTTTAAATAAGATTGGAGTGG	240		
DB	425	TTTATACCCAGCTCTCCATACAAATAGGTGCGAATTTTAAATAAGATTGGAGTGG	484		
QY	241	ATGGACATTTTGTACCATGAGATTGGGACACGCGAGTCCAGAGCGATCCGAGGCC	300		
DB	485	WTTKACATTTTGTACCATGAGATTGGGACACGCGAGTCCAGAGCGATCCGAGGCC	544		
QY	301	TGAGGACACATTTTACAGAGTTCTGAGTCTGCTGCTTACTTTTATGTCGATGAT	360		
DB	545	TGATKACACATTTTACAGAGTTCTGAGTCTGCTGCTTACTTTTATGTCGATGAT	604		
QY	361	CACAAAGCTTCAGAACTTAAGTAACTGAGAAAGAAATTCATATATTATGCAATGTA	420		
DB	605	CACAAAGCTTCAGAACTTAAGTAACTGAGAAAGAAATTCATATATTATGCAATGTA	664		
QY	421	AAGAGCTGAGAGCTTTCTTTTGTGATTCTGGGTAAACAAGATTGACATAAGCGGCG	480		
DB	665	AAGAGCTGAGAGCTTTCTTTTGTGATTCTGGGTAAACAAGATTGACATAAGCGGCG	724		
QY	481	AGGTGCTACAGAAAGCCGAGCTTGTGTCAGGAGCAACGGCGACATCTTATTTTG	540		
DB	725	AGGTGCTACAGAAAGCCGAGCTTGTGTCAGGAGCAACGGCGACATCTTATTTTG	784		
QY	541	AAACAAGTGCAAAAGATGCCAAATGTGGCAGCAGCTTTGAGGAAGCGGTTCGAAG	600		
DB	785	AAACAAGTGCAAAAGATGCCAAATGTGGCAGCAGCTTTGAGGAAGCGGTTCGAAG	844		
QY	601	TTCTTGCTACCGAGATAGGTGAGATCAATTTGATTTCAGACAGACACATCTTACC	660		
DB	845	TTCTTGCTACCGAGATAGGTGAGATCAATTTGATTTCAGACAGACACATCTTACC	904		
QY	661	GAAAGCCCAAGCTAGCTCATCTTCTGTTGATTGTTAGATTGTTGATCTTCAACA	720		
DB	905	GAAAGCCCAAGCTAGCTCATCTTCTGTTGATTGTTAGATTGTTGATCTTCAACA	964		
QY	721	ACTCACATATACACAAATCAACATGGGATGGAGAGAAATAGCGTTTCAGCAG	780		
DB	965	ACTCACATATACACAAATCAACATGGGATGGAGAGAAATAGCGTTTCAGCAG	1024		
QY	781	TGT 783			
DB	1025	TGT 1027			

RESULT 11

BQ772388/c

LOCUS

DEFINITION UI-H-F80-bbq-e-19-0-UI.e1 NCI-CGAP\_F80 Homo sapiens cDNA clone

UI-H-F80-bbq-e-19-0-UI 3', mRNA sequence.

720 bp mRNA linear EST 26-JUL-2002

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BQ772388  
BQ772388.1 GI:21980864  
Homo sapiens (human)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 720)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)  
Tissue Procurement: James Martin  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
Seq primer: M13 FORWARD  
POLYA-Yes.

FEATURES  
source

Location/Qualifiers  
1..720  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-F80-bbq-e-19-0-UI"  
/tissue\_type="Chondrosarcoma Cell line"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP\_F80"  
/notes="Vector: pT73-Pac (Pharmacia) with a modified polylinker. Site 1: EcoR I; Site 2: Not I; NCI CGAP\_F80 is a cDNA library containing the following tissue(s): a pool of 3 chondrosarcoma cell lines (grade 2) The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dfr)18 tail. The sequence tag for this library is CGTACGGAC. The cell lines was provided by Dr James Martin of University of Iowa.  
TAG LIB=UI-H-F80  
TAG\_TISSUE=Human grade 2 chondrosarcoma cell line pool  
TAG\_SEQ=CGTACGGAC"

BASE COUNT 216 a 138 c 112 g 254 t

Query Match

Best Local Similarity 59.9%; Score 704.4; DB 13; Length 720;

Matches 708; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY	462	ATTGACATAGCAAGCGGAGGTGTCTACAGAAAGCCCAAGCTTGGTCAGGGACAAC	521
DB	720	ATTGACATAGCAAGCGGAGGTGTCTACAGAAAGCCCAAGCTTGGTCAGGGACAAC	661
QY	522	GGCGATCTCTTATTTTGAACAAAGTGCACAAATGTCAGAGAGGCGCTTT	581
DB	660	GGCGATCTCTTATTTTGAACAAAGTGCACAAATGTCAGAGAGGCGCTTT	601
QY	582	GAGGAAGCGGTTCGAAGAGTCTTGTACCGAGGATAGGTAGATCATTTGATTCAGACA	641
DB	600	GAGGAAGCGGTTCGAAGAGTCTTGTACCGAGGATAGGTAGATCATTTGATTCAGACA	541
QY	642	GACACAGTCAATCTTCCAGAAAGCCCAAGCTTCTCTGCTGTTGTTGTTAGAT	701
DB	540	GACACAGTCAATCTTCCAGAAAGCCCAAGCTTCTCTGCTGTTGTTGTTAGAT	481

QY	702	TGTTGATGC	AATCTTAA	CAAC	CTCA	CACATATAC	ACAAATCA	CAATGGGATGGAGAGA	761
Db	480	TGTTGATGC	AATCTTAA	CAAC	CTCA	CACATATAC	ACAAATCA	CAATGGGATGGAGAGA	421
QY	762	GAATTAGCG	TTTGCAG	CAGTGTAT	CACTACT	CAATAAAT	TAAACTAA	TCTGCTCGCTTC	821
Db	420	GAATTAGCG	TTTGCAG	CAGTGTAT	CACTACT	CAATAAAT	TAAACTAA	TCTGCTCGCTTC	361
QY	822	ATTAGTTGG	TGGGAGA	AGGACAC	CATCCACT	TTTGGAGGA	ATATATTAT	CTCAATATATGG	881
Db	360	ATTAGTTGG	TGGGAGA	AGGACAC	CATCCACT	TTTGGAGGA	ATATATTAT	CTCAATATATGG	301
QY	882	CACCTTAC	ATTATAA	ATCTG	ACAGTTG	CTCAATAC	GGTTCTTTTAA	TTTAAATATATGTA	941
Db	300	CACCTTAC	ATTATAA	ATCTG	ACAGTTG	CTCAATAC	GGTTCTTTTAA	TTTAAATATATGTA	241
QY	942	AGTTGCAG	AGCTAAT	TAAATGAA	TGACCA	AGACTTTAA	TATATAA	TAAAAATAGAAACTT	1001
Db	240	AGTTGCAG	AGCTAAT	TAAATGAA	TGACCA	AGACTTTAA	TATATAA	TAAAAATAGAAACTT	181
QY	1002	GACTATCT	TAGAAGTT	TATAC	TCTGGAT	TTTTTCT	CTGGAAAA	TGGAGAACTACTCTTTTATATA	1061
Db	180	GACTATCT	TAGAAGTT	TATAC	TCTGGAT	TTTTTCT	CTGGAAAA	TGGAGAACTACTCTTTTATATA	121
QY	1062	TGTGTATG	TTTTTATG	CAATAG	CAAT	TGATTTCT	TGTGTT	CAGGGAATACCTTTCCCTAAAG	1121
Db	120	TGTGTATG	TTTTTATG	CAATAG	CAAT	TGATTTCT	TGTGTT	CAGGGAATACCTTTCCCTAAAG	61
QY	1122	CAATAATG	TTAGATAT	TAAAGAT	TAAAACT	TAATGT	TAAAAA	AAAAA	1175
Db	60	CAATAATG	TTAGATAT	TAAAGAT	TAAAACT	TAATGT	TAAAAA	AAAAA	7

```

RESULT 12
CD515129
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CD515129 881 bp mRNA linear EST 06-JUN-2003
AGENCOURT 14365496 NIH MGC 181 Homo sapiens cDNA clone
IMAGE:30398118 5', mRNA sequence.
CD515129
CD515129.1 GI:31446847
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 881)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg 31 Rm10A07 Bethesda, MD 20892
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: ND4W482 row: 1 column: 07
High quality sequence stop: 714.
Location/Qualifiers
1. .881
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30398118"
/tissue_type="White Matter"
/dev_stage="Unknown"
/lab_host="DH10B-Tcon A ( T1 and T5 phase resistances) "
/clone_lib="NIH MGC 181"

```

```

/note="vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV
(destroyed); Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.42 kb. Library was constructed by
(Inventrogen). Note: this is a NIH_MGC Library."
238 a 187 c 227 q 229 t
BASE COUNT

```

[illegible]

RESULT 13	AI479232/c	AI479232	767 bp	linear	EST 14-APR-1999
LOCUS	AI479232	un56805.x1	NCI CGAP Kid11 Homo sapiens	cdna clone IMAGE:2162096.3,	
DEFINITION	AI479232	similar to SW:RAB9_HUMAN P51151 RAS-RELATED PROTEIN RAB-9. ;, mRNA			
ACCESSION	AI479232	sequence.			
VERSION	AI479232.1	GI:4372400			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

FEATURES  
source

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
		Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
		1 (bases 1 to 767)		
		NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .		
		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
		Tumor Gene Index		
		Unpublished		
		Contact: Robert Strausberg, Ph.D.		
		Email: cgapbs-remail.nih.gov		
		Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.		
		Emmert-Buck, M.D., Ph.D.		
		cDNA Library Preparation: M. Bento Soares, Ph.D.		
		cDNA Library Arrayed by: Greg Lennon, Ph.D.		
		DNA Sequencing by: Washington University Genome Sequencing Center		
		Clone distribution: NCI-CCAP clone distribution information can be		
		found through the I.M.A.G.E. Consortium/LLNL at:		
		<a href="http://www.bio.lnl.gov/brp/image/image.html">www.bio.lnl.gov/brp/image/image.html</a>		
		Insert Length: 244 Std Error: 0.00		
		Seq primer: -40UP from Gibco		
		High quality sequence stop: 454.		
FEATURES		Location/Qualifiers		
	source	1..767		
		/organism="Homo sapiens"		
		/mol_type="mRNA"		
		/db_xref="taxon:9606"		
		/clone="IMAGE:2162096"		
		/lab_host="DH10B"		
		/clone_lib="NCI CGAP Kid11"		
		/note="Organ: kidney; Vector: p7T73D-Pac (Pharmacia) with		
		a modified polylinker; Site 1: Not 1; Site 2: Eco RI;		
		Plasmid DNA from the normalized library NCI CGAP Kid3		
		prepared, and ss circles were made in vitro. Following HAP		
		purification, this DNA was used as tracer in a subtractive		
		hybridization reaction. The driver was PCR-amplified cDNAs		
		from a pool of 5,000 clones made from the same library		
		(cloneId 1322376-1323911, 1456007-1456775, and		
		1500552-1502855). Subtraction by Bento Soares and M.		
		Fatima Bonaldo."		
BASE COUNT	213 a	161 c	135 g	254 t
ORIGIN				4 others
Query Match		58.9%	Score 692.2;	DB 9; Length 767;
Best Local Similarity		95.4%	Pred. No. 1.1e-84;	
Matches	73;	Conservative	0;	Mismatches 33; Indels 2; Gaps 2;
Qy	249	TTTGTTACCAATGAGATTTGGACACGGCAGGTCAGGAGCGATTCCGAAGCCTGAGGACA	308	
Db	765	TTTTGTNACATGCAAGATTTGGNACACGACAGGTCAGAGCGATTTCGAAGCCTGAGGACA	706	
Qy	309	CCATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAACTGTCGATGATTCACAAAGC	368	
Db	705	-CATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAACTGTCGATGATGATTCACAAAGC	648	
Qy	369	TTCCAGAACTTAAGTAACTCGAAGAAAGATTTCATATATTATGATGATGGAAGAGCCT	428	
Db	647	TTCCAGAACTTAAGTAACTCGAAGAAAGATTTCATATATTATGATGATGGAAGAGCCT	588	
Qy	429	GAGAGCTTCTTTTGTGATCTTGGGTAAACAAGATTGACATAAGCGAAGCGAGGTGTCT	488	
Db	587	GAGAGCTTCTTTTGTGATCTTGGGTAAACAAGATTGACATAAGCGAAGCGAGGTGTCT	528	
Qy	489	ACAGAAGAGCCCAAGCTTGGTCAGGGACAAAGCGGACTATCTTTATTTTGAACAAGCT	548	
Db	527	ACAGAAGAGCCCAAGCTTGGTCAGGGACAAAGCGGACTATCTTTATTTTGAACAAGCT	468	
Qy	549	GCAAAAGATGCCCAAAATGTCAGCAGCGCTTTGAGGAAGCGGTTTCCGAAGAGTTCTTGCT	608	
Db	467	GCAAAAGATGCCCAAAATGTCAGCAGCGCTTTGAGGAAGCGGTTTCCGAAGAGTTCTTGCT	408	
Qy	609	ACCGAGATAGGTCAGATCATTTGATTCGACAGACACAGTCGATCTTACCAGAAAGCCC	668	
Db	407	ACCGAGATAGGTCAGATCATTTGATTCGACAGACACAGTCGATCTTACCAGAAAGCCC	348	
Qy	669	AAGCCTAGCTCACTTCTGCTGTGATTTGATGATTTGTTAGATTTGTTCAACCAACTCACAC	728	

Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

```
BASE COUNT      214 a 164 c 134 g 261 t      5 others
ORIGIN

Query Match      58.6%; Score 688.6; DB 9; Length 778;
Best Local Similarity 96.5%; Pred. No. 3.3e-84;
Matches 722; Conservative 0; Mismatches 24; Indels 2; Gaps 2;

QY. 267 TGGACACCGCAGGTCAGGAGCGATTCCGAAGCTGAGGACACCAATTTACAGAGTTCT 326
Db 754 TTGGACACCGCAGGTTCCAGGAGCGATT-CCGAAGCTTAGGACACATTTTACAGAGTTCT 696
QY 327 GACTGCTGCTGCTTACTTTTAGTGTGATGATTCCAAAGCTTCCAGACTTAAGTAAC 386
Db 695 GACTGCTGCTGCTTACTTTTAGTGTGATGATTCCAAAGCTTCCAGACTTAAGTAAC 637
QY 387 TGAAGAAAGAAATTCATATATTATGACAGATGTGAAGAGCCTGAGAGCTTTCCTTTGTG 446
Db 636 TGAAGAAAGAAATTCATATATTATGACAGATGTGAAGAGCCTGAGAGCTTTCCTTTGTG 577
QY 447 ATTCTGGGTACAGATTGACATAAGCGACGCGAGTGTCTACAGAGAGCCCAAGCT 506
Db 576 ATTCTGGGTACAGATTGACATAAGCGACGCGAGTGTCTACAGAGAGCCCAAGCT 517
QY 507 TGGTGCAGGGACACACCGGCTATCTCTTATTTTGAACCAAGTGCACCAAAAT 566
Db 516 TGGTGCAGGGACACACCGGCTATCTCTTATTTTGAACCAAGTGCACCAAAAT 457
QY 567 GTGGCAGCAGCTTTGAGGAGCGGTTGGAAGATTTCTGCTACCGAGATAGTGCAGAT 626
Db 456 GTGGCAGCAGCTTTGAGGAGCGGTTGGAAGATTTCTGCTACCGAGATAGTGCAGAT 397
QY 627 CATTTGATTCAGACAGACACAGTCAATCTTCCACGAAAGCCCAAGCTAGCTCATCTGC 686
Db 396 CATTTGATTCAGACAGACACAGTCAATCTTCCACGAAAGCCCAAGCTAGCTCATCTGC 337
QY 687 TGTGATTTAGATTTGTTGATGCTTAACTCAACCACTCACACATATACCAAAATCAACA 746
Db 336 TGTGATTTAGATTTGTTGATGCTTAACTCAACCACTCACACATATACCAAAATCAACA 277
QY 747 TGGGGATGGAGAGAGAAATAGCGTTTTCAGCAGCTGATCATCTACTATATAAATTAAC 806
Db 276 TGGGGATGGAGAGAGAAATAGCGTTTTCAGCAGCTGATCATCTACTATATAAATTAAC 217
QY 807 TAATGTTGCTGCTTCAATAGTTGGTGGAGAGGGACACATCCACTCTTTGGAGGAATATA 866
Db 216 TAATGTTGCTGCTTCAATAGTTGGTGGAGAGGGACACATCCACTCTTTGGAGGAATATA 157
QY 867 TTTACTCAATAAGGCACCTTACATTTATATAATTTGAACAGTTGCTATATAAGTTTCTT 926
Db 156 TTTACTCAATAAGGCACCTTACATTTATATAATTTGAACAGTTGCTATATAAGTTTCTT 97
QY 927 TAATTTAAATATGTAAGTTGACAGCTAATAAATGAATGACCAAGCTTTAATTAAT 986
Db 96 TAATTTAAATATGTAAGTTGACAGCTAATAAATGAATGACCAAGCTTTAATTAAT 37
QY 987 AAAAAAAGAACTTCACTATTTCTAGAA 1014
Db 36 AAAAAAAGAACTTCACTATTTCTAAAA 9

RESULT 15
AI758228/c
LOCUS
DEFINITION
Cy6b09.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:2284025 3,
similar to SW:RAB9_HUMAN P51151 RAS-RELATED PROTEIN RAB-9. 1, mRNA
sequence.
ACCESSION
AI758228
VERSION
AI758228.1 GI:5151951
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 708)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 450.  
Location/Qualifiers  
1..708  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="IMAGE:2284025"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Kid11"  
/notes="Organ: kidney; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 202 a 144 c 125 g 234 t 3 others  
ORIGIN

Query Match 58.4%; Score 686; DB 9; Length 708;  
Best Local Similarity 98.7%; Pred. No. 7.8e-84;  
Matches 700; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

```
QY 300 CTGAGGACACCATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAGTTCGATGAT 359
Db 708 CTTAGGACACCATTTTACAGAGTTCTGACTGCTGCTGCTTACTTTTAGTTCGATGAT 650
QY 360 TCACAAAGCTTCAGAACTTAAGTAACCTGGAAGAAAGATTCATATATTATGAGATGTG 419
Db 649 TCACAAAGCTTCAGAACTTAAGTAACCTGGAAGAAAGATTCATATATTATGAGATGTG 590
QY 420 AAAGAGCTGAGAGCTTTCTTTTGTGATTTCTGGTTAAAGATTCACATAAGCGACGG 479
Db 589 AAAGAGCTGAGAGCTTTCTTTTGTGATTTCTGGTTAAAGATTCACATAAGCGACGG 530
QY 480 CAGGTGCTACAGAAAGCCCAAGCTTGTGTCAGGACCAACCGCGACTATCTTATTTT 539
Db 529 CAGGTGCTACAGAAAGCCCAAGCTTGTGTCAGGACCAACCGCGACTATCTTATTTT 470
QY 540 GAACAAAGTGCACAAAGATGCCACAAATGTGGCAGAGCCTTTGAGGAGCGGTTGAGAA 599
Db 469 GAACAAAGTGCACAAAGATGCCACAAATGTGGCAGAGCCTTTGAGGAGCGGTTGAGAA 410
QY 600 GTTCTTGTACCGAGGATAGGTGAGATTCATTTGATTCAGACACACAGTCAATCTTAC 659
Db 409 GTTCTTGTACCGAGGATAGGTGAGATTCATTTGATTCAGACACACAGTCAATCTTAC 350
QY 660 CGAAGCCCAAGCTAGCTCATCTTCTGTTGATTTGTTAGATTGTTGATTCATTAACC 719
Db 349 CGAAGCCCAAGCTAGCTCATCTTCTGTTGATTTGTTAGATTGTTGATTCATTAACC 290
QY 720 AACTCACATATACACAAAATCAACATGGGATGGAGAGAAATTAGCGTTTGACGCA 779
```

